U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

MSc

MSc

MSc

MSc

MSc

2.º
CICLO

FCUP
2014

U.PORTO

Phylogeography of mtDNA haplogroup L2

Marina Soares da Silva

FC

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

25 ANOS
A DESCOBRIR O CANCRO
ipatimup

Phylogeography of mtDNA haplogroup L2

Marina Soares da Silva

Dissertação de Mestrado apresentada à
Faculdade de Ciências da Universidade do Porto

Mestrado em Biodiversidade, Genética e Evolução

2013 / 2014

Phylogeography of mtDNA haplogroup L2

Marina Soares da Silva
Mestrado em Biodiversidade, Genética e Evolução
Departamento de Biologia
2013/2014

Orientador
Pedro Alexandre Dias Soares, PhD, Investigador no Instituto de Patologia e Imunologia Molecular da Universidade do Porto (Ipatimup)

Coorientadora
Luísa Maria de Sousa Mesquita Pereira, PhD, Investigadora no Instituto de Patologia e Imunologia Molecular da Universidade do Porto (Ipatimup)

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____

Dissertação de candidatura ao grau de Mestre em Biodiversidade, Genética e Evolução submetida à Faculdade de Ciências da Universidade do Porto.

O presente trabalho foi desenvolvido no Instituto de Patologia e Imunologia Molecular da Universidade do Porto (Ipatimup), sob a orientação científica do Doutor Pedro Alexandre Dias Soares e coorientação da Doutora Luísa Maria de Sousa Mesquita Pereira.

Dissertation for applying to a Master's degree in Biodiversity, Genetics and Evolution, submitted to the Faculty of Sciences of the University of Porto.

The present work was developed at the Institute of Molecular Pathology and Immunology of the University of Porto (Ipatimup), under the scientific supervision of Pedro Alexandre Dias Soares, PhD and co-supervision of Luísa Maria de Sousa Mesquita Pereira, PhD.

“For Africa to me... is more than a glamorous fact. It is a historical truth. No man can know where he is going unless he knows exactly where he has been and exactly how he arrived at his present place.”

Maya Angelou

Agradecimentos

Em primeiro lugar quero agradecer ao Pedro, meu orientador, por tudo o que me ensinou este ano, por mesmo longe ter continuado presente e por ter sido incansável nesta "maratona" final. Mas acima de tudo, obrigada por nunca me teres cortado as asas e teres feito com que eu sinta que este projeto me pertence.

À Luísa, agradeço por me ter aceitado no grupo e pela coorientação neste trabalho. Muito obrigada por esta oportunidade que me permitiu aprender tanto.

Agradeço a todo o grupo de Diversidade Genética e restantes Ipatimupianos que me receberam e foram uma preciosa ajuda ao longo deste ano. Um obrigada especial à Mafalda e ao Rob, por terem estado sempre disponíveis e por tantas vezes terem transformado as 15h30 em 17h no serviço de sequenciação, e ao Bruno, cuja ajuda foi sempre essencial para resolver todos os dramas informáticos.

Obrigada às meninas Andreia, Joana, Marisa e Sofia, por todo o apoio, injeção de confiança e companhia. Obrigada pela ajuda em todos os momentos críticos e por todas as horas de almoço, cafés e lanches em que viajámos pelo mundo fora (até Westeros). E um obrigada particular ao Joel, que já desde o tempo de estágio me acompanha nestas andanças.

À turma de 2012/2013 de MBGE, obrigada pelo ano em Vairão e, claro, parabéns a todos! Um obrigada à Joana por ter sido a minha companhia em todos os trabalhos e uma ajuda essencial em muitos momentos. (Conseguimos!!!)

Aos meus amigos de sempre, malta do costume e capas negras,… mais do que um obrigada, fica um pedido desculpa pela ausência ao longo deste ano. Não estão esquecidos!

Ao Carlos… não quero cair nos clichés habituais, não há nada para dizer que não tenha já sido dito. És a pessoa que melhor compreende tudo "isto". Fazemos uma equipa incrível! Um obrigada do tamanho do mundo que um dia havemos de conhecer juntos.

Aos meus pais, muito obrigada por me terem dado a oportunidade de estudar, um privilégio que foi negado à maioria da nossa família. Tentei estar à altura, espero encher-vos sempre de orgulho. Dedico esta tese aos meus avós, que sempre me ensinaram a importância de conhecermos e orgulharmo-nos das nossas origens. São a prova viva de que do (muito) pouco é possível fazer imenso. Sou uma sortuda por vos ter! Muito obrigada!

Resumo

O haplogrupo de DNA mitocondrial (mtDNA) L2 teve origem no Oeste de África, mas encontra-se atualmente disperso por todo o continente, sendo também bastante frequente no Este e Sul de África. Movimentos associados ao haplogrupo L2 foram anteriormente descritos como estando relacionados com a expansão Bantu, que atravessou a África subsariana nos últimos 5 mil anos (5 ka). Contudo, na região do Este de África onde a frequência descrita de L2 é a mais alta (superior a 30 % na região do Sudão e Etiópia) não existem registos de grupos Bantu. Adicionalmente, uma análise realizada anteriormente baseada em HVRI (Região Hipervariável I) da molécula de mtDNA mostrou que a expansão do L2 para o Este deverá ter ocorrido muito mais cedo, potenciada pela melhoria das condições climáticas do início do Holoceno (há aproximadamente 10-12 ka). O presente estudo pretende reconstruir a filogenia do haplogrupo L2 de forma a fornecer informação sobre a complexa teia de migrações que ocorreram em África nos últimos milhares de anos.

Sequenciámos 44 mitogenomas representativos da diversidade do haplogrupo em África, com principal foco no Este de África (nomeadamente Etiópia, Somália e Sudão/Sudão do Sul), uma região pouco representada na maioria dos estudos. Estas sequências foram combinadas com mais de 750 sequências completas de L2 disponíveis em bases de dados *online*. A reconstrução filogenética foi levada a cabo utilizando Máxima Parcimónia, Máxima Verosimilhança e inferência Bayesiana, aplicando diferentes relógios moleculares (incluindo um relógio aplicado à molécula completa corrigido para seleção purificadora, um relógio para mutações sinónimas e um relógio relaxado). Para fins comparativos, efetuou-se a análise filogenética de outra linhagem mitocondrial, haplogrupo L0a (baseada em sequências completas publicadas), dado alguns paralelismos encontrados nos padrões filogenéticos associados a grupos Bantu. Este estudo foi complementado com testes populacionais, de forma a avaliar as dinâmicas populacionais entre Bantus e populações locais do Este na região dos Grandes Lagos antes da expansão de falantes de idiomas Bantu em direção ao sul.

A análise filogenética das sequências completas de L2 indica que as linhagens sul-africanas diferem das do Este e agrupam diretamente com as do centro/oeste num período

mais recente da Pré-história. Por outro lado, as linhagens do Este parecem apresentar estimativas de idades mais antigas, indicando que o haplogrupo L2 provavelmente chegou ao Este de África muito antes da expansão Bantu, durante o início do Holoceno, quando as condições climáticas melhoraram e o fluxo génico era provavelmente frequente na África central. O mesmo padrão pode ser observado na filogenia do subhaplogrupo L0a, mas em direção oposta. L0a originou-se no Este (sendo hoje muito comum no Sul) e moveu-se para a África Central também durante a melhoria das condições climáticas. Tal como as linhagens L2, as L0a do Sul não parecem estar relacionadas com as do Este. Inferência Bayesiana (através de *Bayesian Skyline Plots* – BSPs) suporta estes resultados, indicando incrementos no efetivo populacional (Ne) associado ao L2 em diferentes regiões africanas, consistentes com as idades estimadas para a filogenia deste haplogrupo e coincidentes com diferentes migrações.

Adicionalmente, sinais do tráfico transatlântico de escravos são visíveis na filogenia do L2, particularmente em L2a1. A maioria das sequências americanas agrupam com amostras do Oeste/Centro, contudo, com o aumento da resolução dos ramos do sul, em alguns ramos parece haver uma associação entre amostras americanas e a região da África do Sul, onde L2 também é bastante frequente, consistente com os registos históricos da época.

A análise populacional mostra que os grupos Bantu do Sul agrupam entre populações do Oeste e Centro, mas sem evidência de fluxo génico forte entre populações do Este e do Sul. Populações do Este, por seu turno, parecem diferenciadas das restantes populações subsarianas. Apesar dos Hutu (Ruanda) e dos Luhya (Quénia), ambos Bantu, agruparem com o Este, parecem afastados das populações sul-africanas, sugerindo que este contacto terá provavelmente ocorrido posterior ou independentemente da migração Bantu para o Sul. Juntamente com a ausência de haplogrupos típicos do Este (L4, L5, L6, L3h, L3i), estas evidências sugerem que a permanência dos Bantu no Este de África não resultou em contacto intenso com as populações locais e que a população que migrou para o Sul tinha essencialmente ancestralidade no fundo genético mitocondrial da África Central.

Três momentos de expansão independentes estão associadas ao haplogrupo L2: (1) ~25-30 ka, durante o Paleolítico superior, (2) movimentos pós-glaciares, provavelmente através do Sahel e potencialmente impulsionados pelas melhorias climáticas com o início do Holoceno (~11.5 ka) e, mais recentemente (3) a expansão Bantu (<5 ka) que levou as linhagens L2 para o Sul.

Palavras-chave: mtDNA, haplogrupo L2, subhaplogrupo L0a, expansão Bantu, início do Holoceno.

Abstract

Mitochondrial DNA (mtDNA) haplogroup L2 had its origin in Western Africa but is nowadays spread across the entire continent, being very frequent also in Eastern and Southern Africa. L2 movements were previously postulated to be related to the Bantu expansion, which crossed sub-Saharan Africa in the last 5 thousand years (5 ka). However, in the East African region, with the highest reported frequency of L2 (over 30 %, in the area of Sudan and Ethiopia), there are no record of Bantu groups. Moreover, a previous analysis based on HVRI (Hypervariable region I) of mtDNA showed that the L2 expansion to the East might have occurred much earlier, triggered by the improvement of environmental conditions in the early Holocene (around 10 to 12 ka). The present study aims to reconstruct the phylogeny of L2 haplogroup, in order to provide insights on the complex net of migrations that occurred in Africa in the last thousand years.

We sequenced 44 mitogenomes representative of the haplogroup diversity across Africa, with main focus on East Africa (namely Ethiopia, Somalia and Sudan/South Sudan), a region poorly represented in most studies. We combined these sequences with more than 750 mtDNA L2 complete sequences available in online databases. Phylogenetic reconstruction was performed using Maximum Parsimony, Maximum Likelihood and Bayesian inference, employing different molecular clocks (including a whole-genome clock corrected for purifying selection, a synonymous clock and a relaxed clock). For comparative purposes, a phylogenetic analysis of another mtDNA lineage, subhaplogroup L0a (based on published whole-mtDNA sequences) was also performed, since phylogeographic patterns associated to Bantu groups seem to be similar in some branches of both haplogroups. In addition, this study was complemented with population-based tests, in order to assess the dynamics between Bantu and Eastern local populations in the Great Lakes region before the expansion of Bantu-speakers southwards.

Phylogenetic analysis of complete mtDNA L2 sequences indicates that lineages in Southern Africa differ from the ones in East Africa and cluster directly with Central/West African lineages at a very recent time scale. On the other hand, lineages in the East seem to present much older age estimates, suggesting that L2 lineages most probably arrived in the East earlier

than the Bantu expansion, in the Early Holocene, when climate conditions improved and gene flow across Central Africa was probably frequent. The same pattern is observed in the phylogeny of subhaplogroup L0a, but in an opposite direction. L0a had its origin in the East (being today common in the South) and moved to Central Africa also during the improvement of climatic conditions. Similarly to L2, southern L0a lineages do not seem to be related with those in the East. Bayesian Skyline Plots (BSPs) support these results, indicating increments in the effective population size (Ne) associated to L2 in different African regions, consistent with the age estimates for the L2 clades and with different migrations.

Additionally, we found signs of the transatlantic slave trade in the L2 phylogenetic tree, particularly in L2a1. Most of the American sequences cluster with western/central samples, however, with the increased resolution of southern clades, we also discovered in some branches what seems to be an association between American samples and Southern Africa, where L2 is also highly frequent, consistent with historical records.

Complementary population analysis indicates that Southern Bantu groups cluster within Western and Central populations, but there is no strong evidence of mtDNA gene flow between Eastern and Southern populations. Eastern populations, on the other hand, seem to be differentiated from other sub-Saharan populations. Although the Hutu (Rwanda) and the Luhya (Kenya), both Bantu-speakers, cluster with Eastern populations, they are distant from the other Southern Africans, suggesting that this contact most likely occurred after or independently of the Bantu migration southwards. This evidence, together with the absence of typical Eastern haplogroups (L4, L5, L6, L3h, L3i) in the South, suggest that Bantu permanence in East Africa did not result in strong admixture with local populations and the populations that migrated towards the South had almost entirely ancestry in a Central African mtDNA gene pool.

Three different independent major expansions were associated to haplogroup L2: (1) ~25-30 ka, during the upper Palaeolithic, (2) post-glacial movements, probably across the Sahel and potentially driven by climatic improvement during the early Holocene (~11.5 ka) (when L2 lineages arrived in East Africa) and, more recently (3) the Bantu Expansion (<5 ka) that took L2 towards the South.

Keywords: mtDNA, halogroup L2, subhaplogroup L0a, Bantu expansion, early Holocene.

Table of contents

List of tables

List of figures

List of abbreviations

A	Adenine
AMH	Anatomically Modern Humans
Ta	Annealing temperature
AD	*Anno domini*
AP	Arabian Peninsula
bp	Base pair(s)
BEAST	Bayesian Evolutionary Analysis by Sampling Trees
BSP	Bayesian Skyline Plot
BC	Before Christ
ºC	Celsius degrees
CA	Central Africa
CR	Control Region
C	Cytosine
ºN	Degrees North
DRC	Democratic Republic of Congo
DNA	Deoxyribonucleic acid
D-loop	Displacement loop
EA	Eastern Africa
Ne	Effective population size
EUR	Europe
e.g.	*exempli gratia*
F_{ST}	Genetic distances between pairs of populations
G	Guanine
Hg.	Haplogroup
HKY85	Hasegawa, Kishino and Yano, 1995
H-strand	Heavy strand
HVRI	Hypervariable Region I

HVRII	Hypervariable Region II
i.e.	*id est*
indel(s)	Insertion(s)/deletion(s)
LGM	Late Glacial Maximum
L-strand	Light strand
MCMC	Markov Chain Monte Carlo
ML	Maximum Likelihood
MP	Maximum Parsimony
µL	Microliter
mM	Milimolar
mtDNA	Mitochondrial DNA
MDS	Multidimensional scaling
NCBI	National Center for Biotechnology Information
NRY	Non-recombinant part of the Y-chromosome
NA	Northern Africa
nDNA	Nuclear DNA
numt(s)	Nuclear DNA sequence(s) of mitochondrial origin
no.	Number
O_H	Origin of replication of the H-strand
O_L	Origin of replication of the L-strand
OOA	Out-of-Africa
PAML	Phylogenetic Analysis using Maximum Likelihood
PCR	Polymerase Chain Reaction
POLγ	Polymerase γ
rCRS	Revised Cambridge Reference Sequence
®	Registered trademark
RNA	Ribonucleic acid
N	Sample size
STP	São Tomé and Príncipe
SA	Southern Africa
TAS	Termination-associated sequences
kb	Thousand base pairs
ka	Thousand years (ago)

TMRCA	Time to the Most Recent Common Ancestor
™	Trademark
UV	Ultraviolet
USA	United States of America
U	Uracile
v.	Version
WA	Western Africa

1. Introduction

1.1. Africa: human origin and initial migrations

With over 2000 ethnolinguistic groups that represent nearly a third of the world's languages and a wide variety of subsistence modes [1], Africa encompasses a great cultural diversity that has always fascinated anthropologists. Moreover, the complex demographical history of African populations, together with great variations in climate, diet and exposure to diseases led to the great genetic and phenotypic diversities observed today. Our species has been living in Africa longer than in any other region of the world, inhabiting a wide range of environments (deserts, savannahs, tropical forests, swamps and mountains) and practicing distinct subsistence modes (hunting-gathering, agricultural and pastoralism) [2].

Africa has been considered the cradle of mankind for a long time. Both genetic (uniparental genetic markers and genome-wide diversity indices) and fossil evidence indicate that anatomically modern humans (AMH) originated in this continent probably around 200 thousand years ago (ka) [3], having later spread all over the globe. However, there is still a vigorous debate not only on the specific region within Africa where modern humans first appeared, but also regarding the initial migrations within this continent [4, 5].

Fossil evidence seems to indicate that modern human origins were in East Africa [3, 6], even though some old fossil specimens were found in Morocco and Israel [7, 8]. Genetics, on the other hand, has failed so far in clarifying this question, since different approaches have provided distinct results. Autosomal studies seem to point a southern origin for AMH [9, 10], whereas uniparental markers provide opposite results: Y-chromosome indicates a central-western origin [11], while mitochondrial DNA (mtDNA) diversity analysis suggests an east-central origin [5].

Nevertheless, ancestral African populations are thought to have been genetically structured for more than 100 ka, while humans were still geographically restricted to Africa, before the Out-of-Africa (OOA) migration [4, 12, 13].

1.1.1. Western and Central Africa

Fossil record in Western Africa (WA) is scarce, but there are evidences that AMH were present in the north-western region at least around 40 ka [14, 15]. WA has a complex history of population expansions. Important movements across the Sahel corridor are thought to have occurred ~9 ka [16], when the Sahara desert reached its maximum humidity [17]. The expansion of Bantu-speakers, which had major impact both culturally and genetically, is believed to have started in Central Africa (CA), in the Grassfields region between southeast Nigeria and western Cameroon no longer than 5 ka [16, 18].

More recently, many kingdoms emerged, such as the Ghana Empire (between Niger and Senegal), which arose in the 8th century and is thought to be the oldest occidental African kingdom, known at that time as the “land of the gold”, due to the commerce of gold, especially with the Arab World [16]. The contact of western native populations with the Berbers dates back at least to the 9th century. In the 12th and 13th centuries, the Ghana Kingdom was followed by other empires such as Mali (which conquered part of Ghana territory) and Songhai [16]. During the Age of Discovery, contact with European cultures led to the fall of many of these kingdoms, which were incorporated as colonies of European empires (e.g.: Portuguese, Spanish, French and English). This continuous network of migrations and invasions led to the admixture of peoples with different ancestries.

Nowadays, WA is one of the most extensively sampled areas in Africa. For a long time the Yoruba (from Nigeria) was the only African population included in the HapMap project [19] and considered as the African comparison reference in many studies [20–23]. More recently, the 1000 Genomes project [24] increased the number of African populations sampled, however, they all belong to the same linguistic family (Niger-Kordofanian), and even the only eastern group included (the Luhya from Kenya) is a Bantu-speaking population that shares ancestry with west/central populations [9, 25]. Therefore, a major effort is still needed to fully characterise the genetic African diversity.

On the other hand, since both genetic and historical evidence suggest that most of the ancestry of African Americans derives from WA [9, 26, 27], studies on the genetics of west African populations have been proved useful to detect genetic variants potentially connected to certain diseases in African American populations [25, 28–30].

1.1.2. Eastern Africa

Eastern Africa (EA) not only harbours populations with very ancient ethnic and genetic diversity (with some of the deepest branches of mtDNA and Y-chromosome) [31–35] but it is also considered the most probable origin for the OOA migration, around 70 ka [31–38]. From a linguistic perspective, all the four major African linguistic groups – Afro-asiatic, Khoisan, Niger-Kordofanian (also known as Niger-Congo) and Nilo-Saharan [39] – are present in East Africa.

Ethiopia is one of "the most complex ethnolinguistic region in the continent" [40] and is considered an anthropological reference: it is the home of many hominin fossils, including the earliest remains of AMH (Omo 1, dating around 195 ka) [3]; it has more idioms than all of Europe [41] and its capital, Addis Ababa, is considered the starting point for modern human expansions and used as a reference for neutral genetic diversity measurements [42]. In fact, most of the information available regarding East Africa focuses essentially on Ethiopia and on the Bantu groups in the Great Lakes region, and little is known about the other eastern African countries.

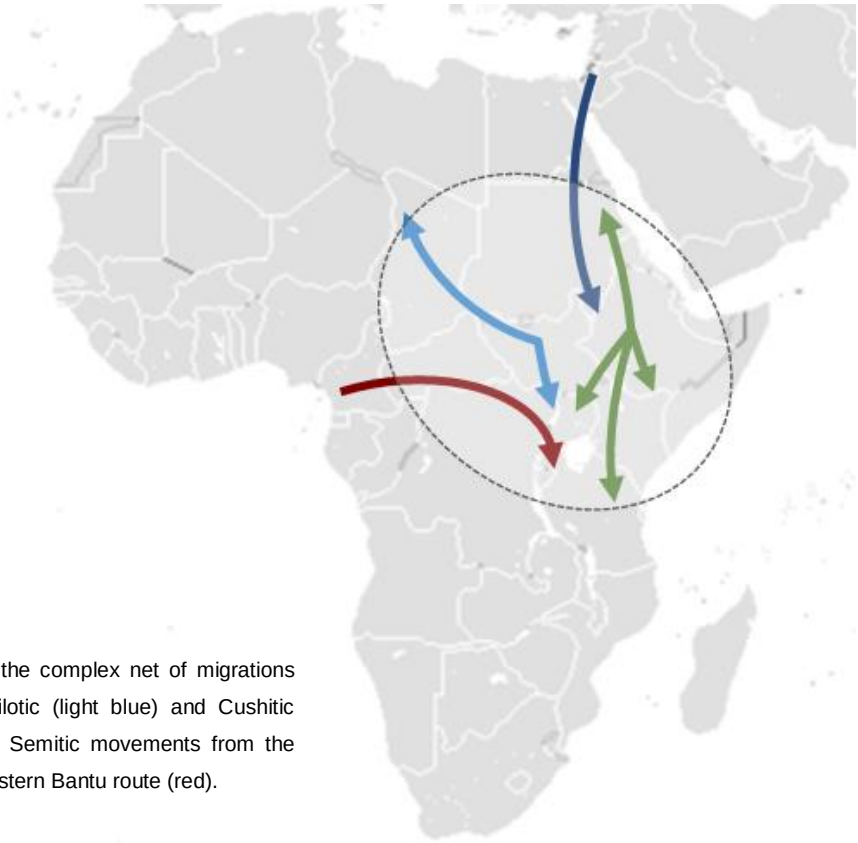

Figure 1. Map representing the complex net of migrations that crossed East Africa: Nilotic (light blue) and Cushitic (green) language dispersals, Semitic movements from the Levant (dark blue) and the eastern Bantu route (red).

East Africa received the influence of various migrations that shaped the current gene pool of the region, either from within Africa or from outside (Figure 1). Two examples are the Nilotic

and Cushitic language dispersals. The Cushitic agropastoralists expanded from Ethiopian highlands into Kenya, Tanzania and into the Red Coast of Sudan (where the Beja people, Cushitic's descendants, live nowadays) [28]. Together with Sudanic hunter-gatherers these agropastoralists were probably responsible for replacing the pre-existing San and Pygmy groups, with their arrival in the proximities of Kenya ~3 ka [40, 43]. The Nilotes, pastoralists who speak Nilo-Saharan idioms, migrated from South Sudan westwards and eastwards by the end of the 1st millenium [28].

There is also evidence of close contact with North Africa. Historical records show that Ethiopia has maintained contact with populations from Egypt across the centuries, mainly due to the trade of myrrh that occurred along the Red Sea coast and dates back to the 3rd millenium [35]. Moreover, genetic data [44] indicate gene flow across the Nile River Valley, between Egypt and South Sudan. Additionally, the Bantu Expansion also left its imprint in the southern part of East Africa, having crossed South Sudan, Uganda, Kenya and Tanzania [45, 46].

There were also movements from the Levant and Arabian Peninsula into East Africa, via the Red Sea [47, 48]. The Horn of Africa is thought to have been the entry point of zebu cattle from the Arabian Peninsula [49] and gene flow between Ethiopia and the Levant, that started around 3 ka, was maintained in the following centuries [50]. Maritime migrations in the region are described at least as early as 8-6 ka, testified by the settlement of Soquotra island [51]. Starting in the 7th millennium BC, due to empire expansions and trade routes (essentially of obsidian and slaves, majority female), gene flow between East Africa and Yemen is thought to have occurred continuously [52, 53], having left its imprint on both African and Arabian populations. Today's Ethiopian Semitic and Cushitic populations are closest to Yemen [50] and the general Yemeni Ethiopian population also shows both cultural and genetic signs of contact with Semitic groups, which occurred in different phases, since the 1st millennium BC [54]. In parallel, the Arabian mtDNA pool contains 6-25 % of African descending lineages, which is consistent with the historical patterns of slave trade in the region [55].

Although several studies have been focusing recently on the genetics of East Africa, its anthropological complexity is still not completely surveyed. This is especially true for small ethnic groups residing in areas enduring military conflicts and social crisis in the past decades, being the ethnic conflict in South Sudan the most recent event. As mentioned above, most information on East Africa (except from the literature about the Bantu expansion) focuses on Ethiopia.

1.1.3. Southern Africa

Nowadays, two main population groups inhabit Southern Africa (SA): the Bantu-speakers and the so-called “Khoisan” groups. However, none of them are thought to have originated in this region.

The Bantu-speakers are believed to have arrived to Angola and Mozambique ~3.5 ka and ~1.8 ka, respectively [56], during a great historical population movement that started in CA and crossed sub-Saharan Africa, known as the Bantu expansion (*see 1.1.5*).

The ‘Khoisan’, on the other hand, (term often loosely applied) are the people from south-western Africa who speak the so-called ‘click languages’ and inhabit the Kalahari desert. However, this group is not phenotypically nor culturally homogeneous [57]. Linguistically, Khoisan represents a complex of three distinct independent language families, spoken both by foragers (self-identified as ‘San’) and pastoralists (the ‘Khwe’), that share the uniqueness of click consonants abundance. Hadza and Sandawe, two linguistic isolates of Tanzania, are also included within the Khoisan group and also feature click consonants [39]. The San descend from autochthonous hunter-gatherers from CA, probably confined to the South due to the advance of Bantu-speakers [16], while an emigration of pastoralists from eastern Africa was the most probable origin of the Khwe [58].

The genetic landscape of SA is incredibly rich, being the region of the world with the highest genetic diversity [59]. According to genome-wide data, the Khoisan hunter-gatherer people of southern Africa are the most genetically diverse of all the human populations [60]. The level of diversity is such that there are more genetic differences between two randomly chosen hunter-gatherer Khoisan individuals than between an European and an Asian [61]. As previously mentioned, due to the high levels of genetic diversity, SA was proposed as the origin for AMH [10].

1.1.4. Climate oscillations

Climatologic evidence suggest that some episodes of climate oscillation might have had major influence on population migrations within Africa and in shaping their structure both before and after the OOA [62–66].

The African Late Glacial Maximum (LGM) might have occurred slightly later than the global LGM [17], around 17-15 radiocarbon (^{14}C) ka (~18-16 calibrated ka), in contrast to the global LGM, 22–19 ka) [67]. Its major feature was aridity, resulting in the expansion of the Sahara several kilometres southwards [17].

The Pleistocene/Holocene transition (~11.5 ka) (Figure 2a) is characterized by changes in atmospheric circulation and solar radiation in the North Hemisphere [64]. During the Holocene climatic optimum (~9-6 ka) north and central African climate was moister than nowadays (Figure 2b). The expansion of rainforest in equatorial Africa, combined with the movement of monsoon rains towards north resulted in the greening of the Sahara [17] and led to the Holocene lacustrine episode, which had its maximum around ~11.3-8.8 ka. During this period the Chad lake was seven times larger than today and the Sahara desert virtually disappeared [68]. Semi-arid and seasonal vegetation covered the Sahara [69] and the Sahel reached as far north as 23 ºN (in contrast with 18 ºN nowadays) [70].

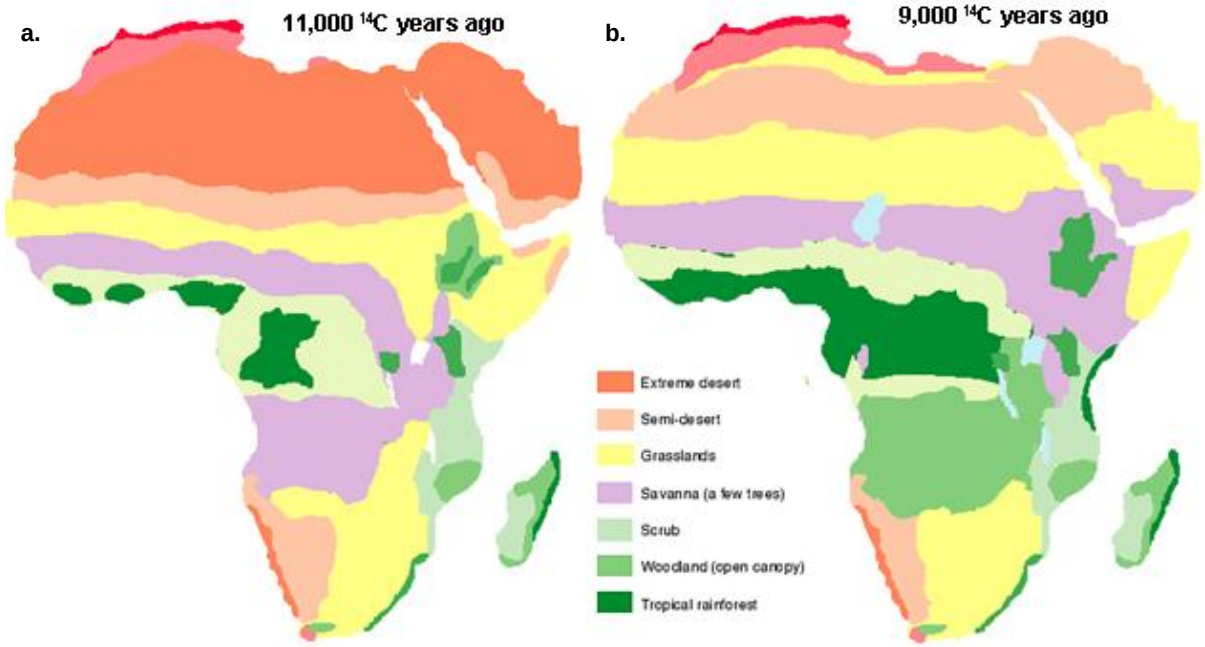

Figure 2. Palaeovegetation in Africa during the Pleistocene/Holocene transition, 11 ^{14}C ka (~12 ka) (a) and during the Holocene climatic optimum 9 ^{14}C ka (~10 ka) (b). Adapted from [17].

Archaeological evidence of human occupation in western Sahara during the late aridity maximum is scarce. During the early Holocene, with the improvement of environmental conditions, major human expansions are thought to have occurred in southwest Asia [71] and in Europe [72]. In Africa, during the early Holocene humid episode, human populations recolonized Saharan areas [64] and gene flow across west/central Africa was probably frequent [4].

1.1.5. The Bantu Expansion

The Bantu language family (included in the Niger-Kordofanian group) is the largest of African languages, with a distribution of approximately 9 million km^2 and 240 million speakers (it is estimated that one African in three speaks a Bantu idiom) [73, 74]. It includes about 500 distinct languages spread across almost half of the African continent: from southern Nigeria in the west to southern Somalia in the east and as far as the Cape [73, 75].

Bantu languages are associated to agricultural lifestyle and iron technology [73] and are believed to have had origin in the Grassfields region between southeast Nigeria and western Cameroon less than 5 ka [18], having later expanded throughout sub-Saharan Africa.

For a long time there was a discussion whether this cultural expansion actually involved movement of people or only diffusion of language and cultural traditions. However, genetic evidence gathered both from autosomes and uniparental markers suggest that the “Bantu expansion” was an actual migration of Bantu-speakers [74]. Some lineages, such as mtDNA haplogroups L0a, L2b, L3b and L3e and Y-chromosome haplogroups E1b1a, E1b1a7 and B2b, were identified as being connected with movements of Bantu populations [4, 32, 76–78].

Figure 3. The Bantu Expansion: (1) western African route and (2) eastern route, depicted against the distribution of the main African language families. Adapted from [79].

The Bantu expansion is thought to have taken two different routes from its starting point, in West-central Africa (Figure 3): (1) a western route, throughout the west coast of Africa, having arrived to Angola, South Africa and Botswana around 3.5 ka, and (2) a eastern route, towards the Great Lakes in East Africa, reaching the region of Uganda about 2.5 ka, where they stayed for a couple thousand years, expanding later into the south, having reached Mozambique by ~1.8 ka [32, 79]. The latter is of particular interest to study potential crossings between migrants and local eastern populations (namely the Nilotic and Cushitic people), during the period in which the Bantu people were stationed in the Great Lakes region.

Linguistic differences between eastern and western Bantu idioms probably reflect these two different routes of expansion. However, a more recent model for the Bantu dispersal assumes a later split of Eastern and Western Bantu [74, 80].

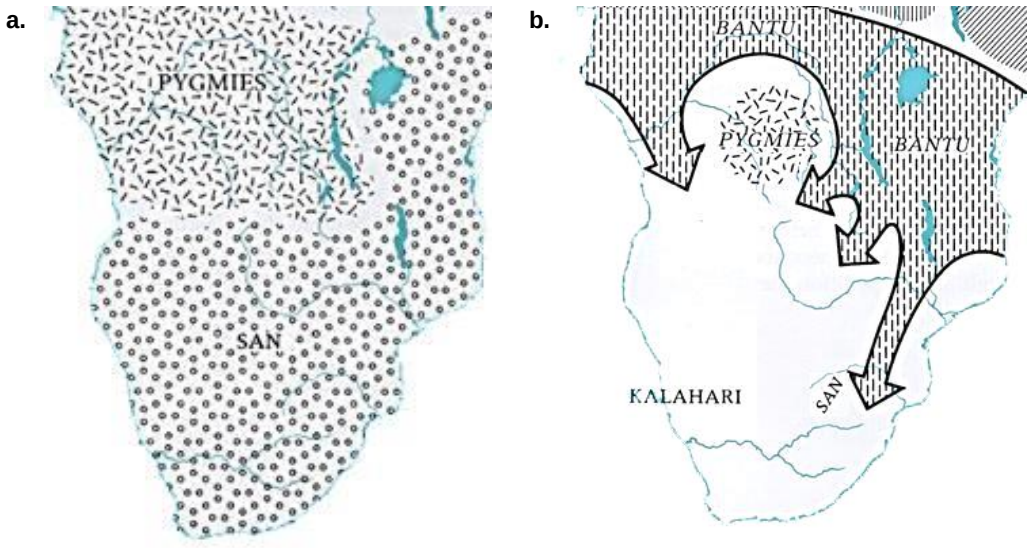

Figure 4. Interaction of Bantu people and other sub-Saharan groups: distribution of Pygmies and San ~8000 BC **(a)** and ~500 AD **(b)**. Adapted from [16].

Either way, the Bantu expansion forced the retreat of contemporary local sub-Saharan populations (Figure 4). The San were further confined to the South, towards the Kalahari desert, and kept their typical Khoisan languages (with click consonants) and ethnic identity. The Pygmies, on the other hand, were pushed deeper into the forests and eventually adopted Bantu languages [16].

1.2. African slavery

On the more recent past, a forced exodus led to the spread of sub-Saharan lineages to different regions of the world. However, the history of African slavery started long before the Age of Discovery. The Arab world has a long history of trading black slaves, which has started in the 7th century and still remains today in Sudan and Mauritania. The main interest of Arab slave trade was in females to serve as domestic servants, entertainers and concubines, whereas males were sought as door-keepers, soldiers or eunuchs [52].

Later, the Portuguese, who had colonies both in Africa and America, were the pioneers of the transatlantic slave trade, but when they first tried to sell black slaves they were not very successful in European markets and most of the slaves were instead resold to North Africa. However, in the first half of the 16th century the traffic of black slaves, mostly male, between the west coast of Africa and its offshore islands increased, as work force for Portuguese sugar plantations at S. Tomé, Madeira and Cape Verde. By the end of the century the trend had already arrived to the American colonies, resulting in a shift in the slave trade, being the New World the main destination [16]. In the beginning of the 18th century the number of slaves shipped across the Atlantic was ten times higher.

1.2.1. Transatlantic slave trade

Although America was probably the last portion of continental land to have been colonized by modern humans and, therefore, the continent with the least human genetic diversity [42], recent historical migrations changed this scenario. In fact, despite the natives having the lowest degree of diversity, America harbours nowadays the most heterogeneous populations in the world. This was the result of ethnic crosses that started in the 15th century between people of different origins: the Europeans colonizers, African slaves and native Amerindians [81].

The Portuguese, who controlled colonies both in America and Africa, were the pioneers of the transatlantic slave trade, but were not the only ones involved. The Spanish and the British also took part in the commerce. Together they were responsible for moving around 11 million African people to America (and at least 2 million more that died during the Atlantic crossing), for more than four centuries (15th-19th centuries). Historical records show that African slaves came

mainly from Western (~8 million slaves) and West-Central (~4 million) Africa and, in a lesser extent, from South-Eastern Africa (~1 million) [26, 27].

Until very recently, both religious and legal restrictions created obstacles against inter-ethnic marriages. Nevertheless, not only do historical records show that inter-ethnic relationships, though stigmatized, were common [82], but there is also genetic evidence of this admixture. There was, however, differential contributions of European and African gender contributions in African American population: elevated European male ancestry, against a high African female contribution [83] (Figure 5).

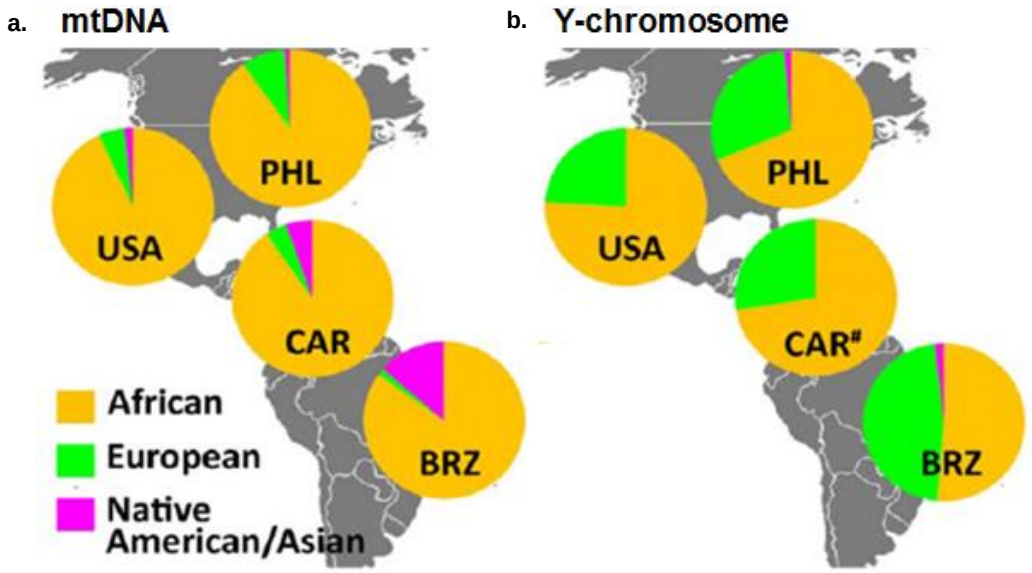

Figure 5. Ancestry of African-descendant Americans: maternal **(a)** and paternal **(b)** components [81].

Regarding the only Portuguese colony in America, Brazil harbours the most heterogeneous populations in the world. Apart from the native indigenous background, Portuguese colonizers and African slaves, Brazil also received more European (mostly from Italia, Spain and Germany) and, in the 20th century, Asiatic immigration [84].

1.3. Reconstructing the past

1.3.1. Archaeogenetics

The term “Archaeogenetics” was coined in 1999, by Amorim [85] and refers to the employment of molecular genetics in order to determine aspects of human prehistory. This

approach is of utmost importance on anthropological studies, since the fossil record of the emergence of our species is scarce and discontinuous.

The first works that aimed to combine genetics and archaeology were, however, conducted much earlier, by Cavalli-Sforza's group in the 70s, who reported the diversity of European populations based on protein markers [86]. Later, the invention of polymerase chain reaction (PCR) [87] became a true milestone for genetic studies, allowing many other techniques (PCR-based methods), such as Sanger's cyclesequencing [88, 89].

DNA studies allow the distinction between female and male lineages, by analysing uniparental molecular markers that can eventually display differing histories of males and females in a population. Nowadays, genetics has been proved useful in revealing prehistoric events, such as migrations (e.g.: the OOA) and palaeodemography [38, 72, 73, 90].

1.3.2. Phylogeography

Avise, the father of Phylogeography, coined the term in 1987 [91]. This discipline aims to apply phylogenetic analysis to understand the distribution of intraspecific genetic variation, taking into account a geographical component.

This type of approach allows to "see genes in space and time" [92] by combining geographic information with an appropriate mutation rate to the molecular marker under study, so as to date a given phylogeny.

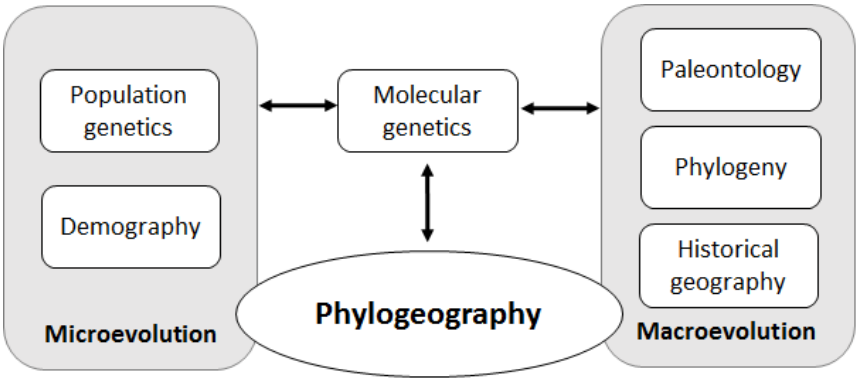

Figure 6. Position of phylogeography within the biodiversity sciences. Adapted from [90].

Phylogeography also combines information from distinct disciplines, typically associated either to the study of microevolution (intraspecific genetic patterns) or of macroevolution (supraspecific variation) (Figure 6). Regarding the study of human evolution, it is important to

enrich phylogeographic studies with data from other fields of research, such as archaeology, linguistics and palaeoclimatology.

1.4. Uniparental genetic markers

Uniparental genetic markers – mtDNA and the non-recombinant part of the Y-chromosome (NRY) – are highly informative in a wide range of fields, including evolutionary anthropology and population history, since they allow to study demographic sex-biased events (female and male-mediated, respectively) due to their uniparental inheritance and lack of recombination, i.e. they allow a reliable reconstruction of the evolution of different lineages based uniquely on mutational events [90]. Whereas Y-chromosome is only present in males, mtDNA is present in both sexes, being transmitted only by females to the following generations (Figure 7).

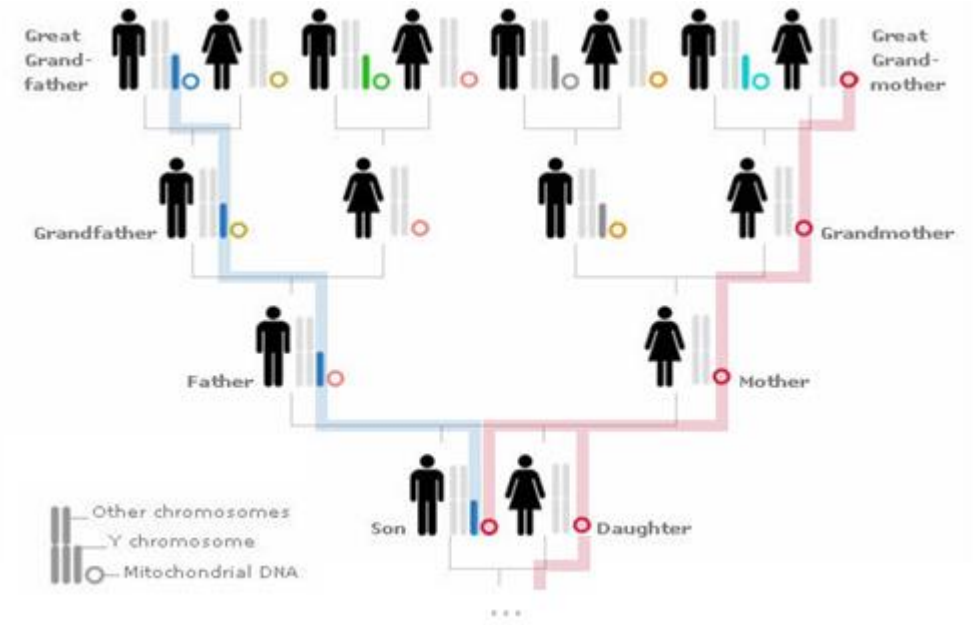

Figure 7. Transmission patterns of uniparental genetic markers (Y chromosome represented in blue and mtDNA in rose). Adapted from The Genographic Project (genographic.nationalgeographic.com).

Uniparental haploid molecular markers allow to frame the demographic events detected within a time period, by including a molecular clock. Therefore, it is possible to infer the age of lineages and expansion times, for instance [90]. Despite slight differences concerning the level of resolution of the phylogenetic tree and coalescent times (inherent to the properties of each

marker), both markers tell a similar demographic history of our species, especially pointing out our African origin [90, 93].

1.4.1. MtDNA

MtDNA, present in mitochondria, is a circular double-stranded molecule, normally as a supercoiled structure, with ~16,569 kilobase pairs (kb) in humans, which codes for proteins involved in the oxidative phosphorylation. The first mitochondrial genome ever sequenced was the human mtDNA genome [94] and it is used as a reference for numbering positions and comparison of polymorphisms in human studies. A few years later the sample was resequenced and the corrected sequence [95], known as the rCRS (revised Cambridge Sequence), is the one used nowadays.

The nucleotide content differs in the two strains, being one richer in Guanine (G), the heavy strand (H-strand), while the other, the light strand (L-strand) has predominately Cytosine (C) [96]. The H-strand is also called leading strand, since it starts replication before the L-strand, also known as lagging strand. There are also differences concerning the amount of genes each chain encodes: the H-strand encodes 28 genes, whereas the L-strand codifies only nine.

MtDNA replication, although controlled by nuclear genes, occurs in the mitochondrial matrix and is independent of nuclear DNA (nDNA) replication or cell phase. It is performed by a specific DNA polymerase, polymerase γ (POLγ) [97]. The exact mechanism of mtDNA replication is still uncertain, but the most accepted model, the asymmetric or displacement model, states that the replication of the two chains has independent and asymmetrical origins [98, 99]. According to this model, replication begins at the O_H site (origin of replication of the H-strand) and is normally “arrested” in the termination-associated sequences (TAS) for transcriptional regulation, forming a triple-stranded structure called the displacement loop (D-loop) of ~700 bp (from the O_H site to the TAS) [99, 100]. Replication of H-strand continues clockwise until exposing the origin of replication of the L-strand (O_L), starting then the replication of the L-strand that proceeds in the opposite direction. The two new strands are then ligated [98, 99].

Another particularity is that the mitochondrial genetic code is slightly different from the nuclear one. UGA is a codon for tryptophan, instead of being a stop codon. The termination

codons in mtDNA transcription are AGG and AGA (arginine codons in the nucleus). AUA codes for methionine, instead of isoleucine [96, 101].

In contrast to nDNA, the mitogenome is highly compact and predominantly coding (with all the protein-coding genes lacking introns), apart from a ~1200 bp segment with regulatory functions, the Control Region (CR) (Figure 8a), extending from position 16,024 to 576 [94]. The CR (includes the D-loop, the O_H and the TAS) contains two hypervariable regions (HVRI and HVRII) (Figure 8b) with higher mutation rates [102]. In the past, sequencing the whole molecule was too difficult or too expensive to do routinely, therefore, most evolutionary studies used to rely on HVRI, due to its rapid accumulation of mutations.

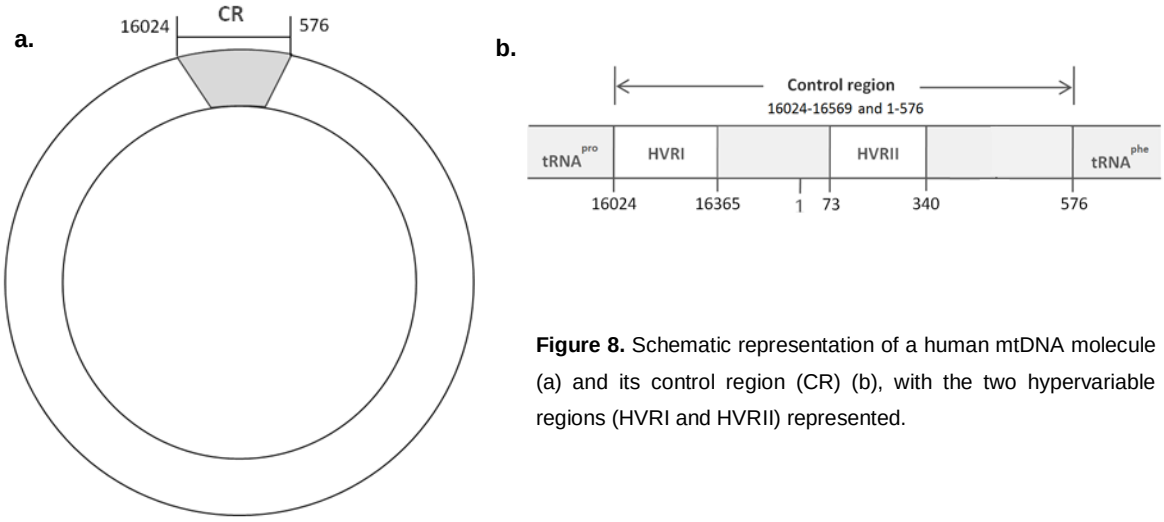

Figure 8. Schematic representation of a human mtDNA molecule (a) and its control region (CR) (b), with the two hypervariable regions (HVRI and HVRII) represented.

Despite the entire molecule not evolving all at the same rate (with the CR, particularly HVRI and HVRII, evolving faster), mtDNA has in general a higher mutation rate than nDNA [103], since it is not protected by proteins, and is highly exposed to radicals of oxygen (resulting from cellular respiration) [96]. The most commonly used mutation rate for human mtDNA is of 1.26×10^{-8} base substitution per site per year [104]. This rate, however, has been questioned in the last years, with the increase of evidence indicating that selective constrains shapes the evolution of mitochondrial genome in humans, as well as in other mammals [37, 105].

An eukaryotic cell has a variable number of mitochondria (a typical mammalian cell contains 1000 to 10,000 mitochondria, depending on the type of cell) and each mitochondrion contains in its matrix several (2-10) copies of mtDNA molecules [106, 107]. Therefore, in well preserved samples, mtDNA is present at a much higher copy number than nDNA. Although it is commonly assumed that all the copies of mtDNA of a given individual are equal (homoplasy), this is not necessarily always the case. Variations might occur between different cells and/or

tissues or even amongst the various mitochondria within the same cell, creating heteroplasmy [96, 102].

During the evolution of eukaryotes, some portions of mtDNA were laterally transferred into nuclear genome at different evolutionary times, originating numts (nuclear DNA sequences of mitochondrial origin) [108, 109]. Once integrated in the nDNA, these fragments slowed down and started evolving at the nuclear mutation rate. Across the human genome there are many numts of various lengths inserted in different chromosomes: between 250 and 600 known [102], but the real number might be higher, since the list available on Mitomap database (www.mitomap.org) was not updated since 2006.

The effective size of mtDNA, exclusively maternally inherited, corresponds to one fourth of autosomal markers, making it more sensitive to demographical events, such as genetic drift, bottlenecks and migrations [102]. Hence, the variation observed in mtDNA sequences is uniquely due to the sequential accumulation of new mutations over generations along maternal lineages. All human mtDNA sequences ultimately trace back to an mtDNA African ancestral lineage (known as the mitochondrial Eve) [111]. The coalescent time to the most recent common ancestor (TMRCA) for human mtDNA is estimated to range ~171−238 ka [37, 112, 113].

Moreover, due to its fast mutation rate, variation accumulates fast enough amongst different geographic locations, making it a suitable molecular marker for phylogeographic approach. MtDNA is a powerful and widely used tool for studying demographic events and migratory occurrences [90, 113–115].

1.4.2. MtDNA phylogeography

Since it is non-recombinant, mtDNA is transmitted as a unique locus, in a block known as haplotype. The term “haplogroup” (Hg) was coined by Torroni in 1993 [116] as a reference to monophyletic clusters of haplotypes, i.e.: all the haplotypes belonging to a given haplogroup share a common ancestor, either extant or reconstructed. Haplogroups are associated with a given geographical distribution, showing great geographic specificity mainly at the continental level.

The first haplogroups defined, baptized A, B, C, and D, were discovered in Native Americans [116]. The haplogroups subsequently described were designated with other letters of

the alphabet. Regarding nomenclature rules, haplogroups are designed with capital letters and more derived subclades are named intercalating lower-case letters with numbers (e.g.: L2, L2a1, L2a1b) [117].

The first attempts to define a global tree for mtDNA variation were based on high-resolution restriction mapping – haplogroups were defined by specific restriction sites – and, despite the low resolution of this approach, the major ancient clades were identified [118, 119]. However, many potential mutations were inevitably not covered and genealogical resolution was not guaranteed. In fact, some early results even supported the Multiregional model [120, 121], in opposition to the OOA model nowadays accepted for mtDNA. Over time, with the employment of novel molecular techniques, such as PCR-based methods, the level of resolution of the tree has increased and the major basal clades were dissected into younger monophyletic branches, representing more restrict geographic and/or ethnic units.

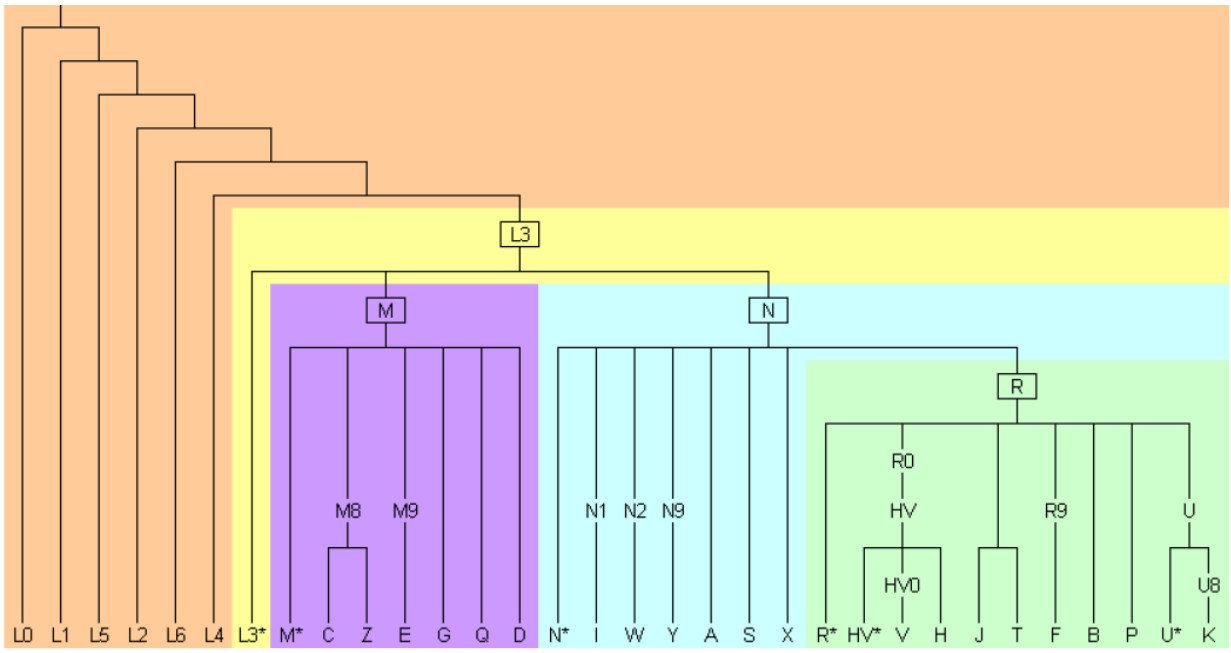

Figure 9. Global mtDNA phylogenetic tree with the main haplogroups represented. L3 harbours all the non African diversity (haplogroups M, N and R – purple, blue and green, respectively) and has some African branches (L3*, in yellow). L* (L0, L1, L2, L4, L5 and L6, in orange) represent all the African maternal variation [122].

From the sequencing of HVRI, a fairly small fragment with only ~340 bp, to the entire CR, and, more recently, to complete sequences, the study of mtDNA as a molecular marker has been on constant evolution. As a result, the resolution of the worldwide mtDNA tree has enormously increased. Nowadays, there are more than 20,600 complete mtDNA sequences available on online databases, such as GenBank® (from NCBI – National Center for Biotechnology Information, Maryland, USA) or 1000 Genomes [24], and the global phylogenetic

tree for human mtDNA (PhyloTree) [122] is constantly updated. Regarding African individuals, more than 2200 mitochondrial genomes are publically available.

The phylogenetic tree of human mtDNA (Figure 9) ultimate traces back to two opposite branches: L0 and L1'2'3'4'5'6 (L1-L6), being the latter more widespread, encompassing almost every mtDNA lineages found nowadays [123]. The mtDNA tree reflects our African origin: not only are the L haplogroups (typically African lineages) the most basal clades of the tree, but also all the non-African mtDNA diversity (macrohaplogroups M, N and R) descends from one specific African haplogroup (L3) [90, 122].

The typology of mtDNA global phylogenetic tree reflects a rapid population growth that seems to have started between 40 to 50 ka (probably corresponding to the cultural revolution of the late Palaeolithic), particularly visible in the lineages that descend from L3 [124, 125], which have long tips. This growth intensified around 10 ka, with the Neolithic revolution (Figure 10). The number of haplogroups also starts increasing almost exponentially around this period of time [126] and many star-like nodes are visible in the global mtDNA tree.

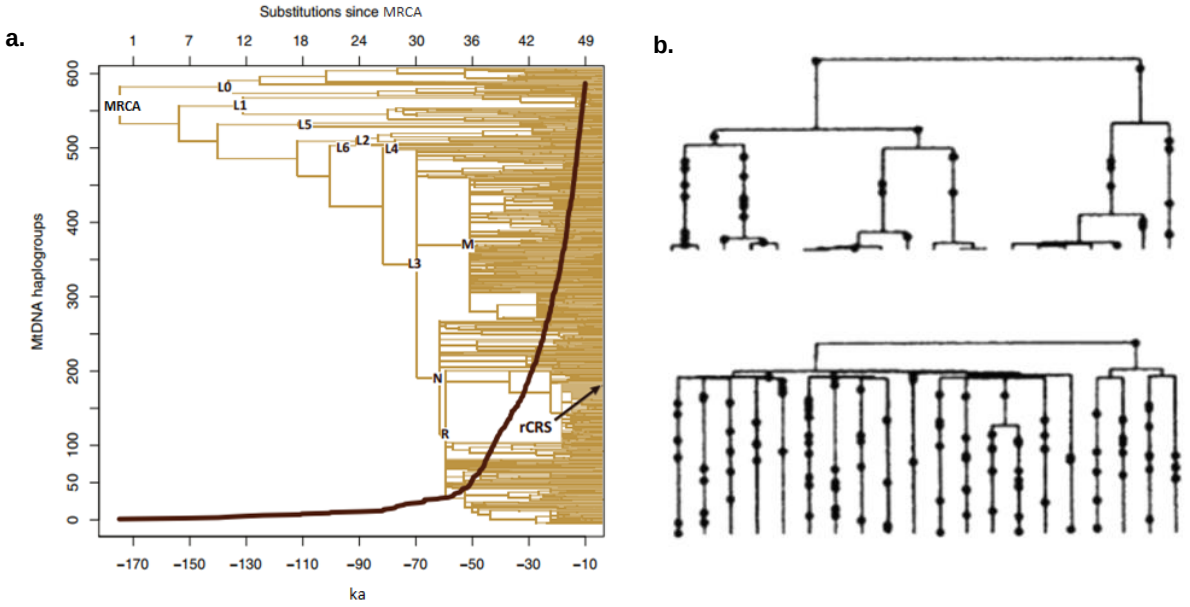

Figure 10. Schematic human mtDNA phylogeny and explosion of haplogroups ~40 ka (a), adapted from [126]. Typical tree topology of population stability, corresponding to the mtDNA before ~40 ka (above) and typical typology of population growth, observed in the human mtDNA, since the late Palaeolithic, ~40ka (under) [125].

The younger lineages, in the tips of the phylogeny, show a higher proportion of nonsynonymous mutations in protein-coding genes and RNA genes, whereas the more ancient branches of the mtDNA phylogeny display a higher proportion of synonymous mutations [127]. This was proved to be evidence of purifying selection acting on mitochondrial DNA, which over

time leads to the loss of lineages containing nonsynonymous potentially deleterious mutations [37]. Therefore, assuming a linear coding-region molecular clock might result in the overestimation of the coalescent times near the tips of the tree.

1.4.3. Haplogroup L2

Haplogroup L2 is the sister branch of the Eastern African L3'4'6 clade that contains all the OOA diversity. While L3'4'6 was originated in East Africa, the mtDNA haplogroup L2 probably originated in West Africa and is widely spread in the continent, being one of the most frequent in many regions, such as in West-Central and South-Eastern (probably associated to Bantu expansion that occurred in the last few millennia) and in North-Western, most likely due to trans-Saharan slave trade [32, 128]. Together with L3, it represents ~70 % of sub-Saharan maternal variation, but, in spite of its high frequency and wide distribution, L2 was not involved in the OOA [129], since it was most likely not present in EA by the time of the OOA.

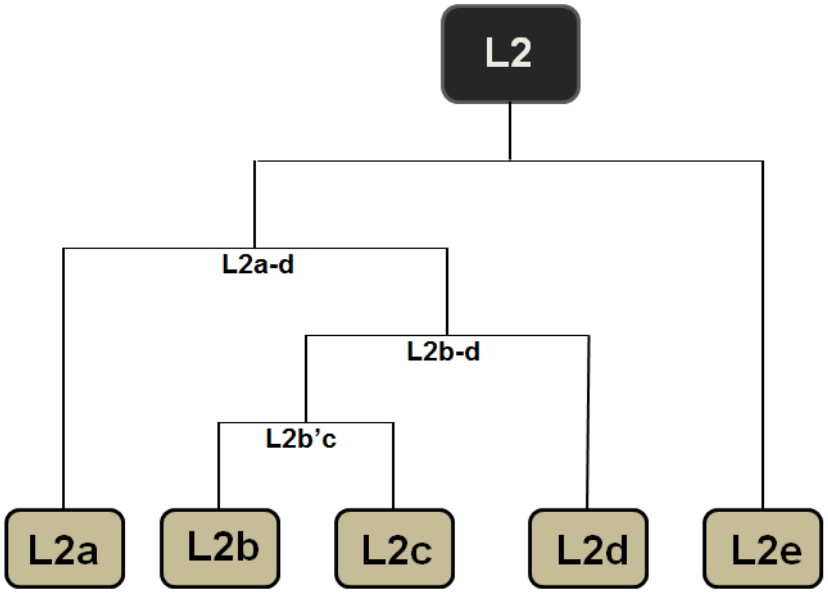

Figure 11. Schematic representation of the five main branches of mtDNA haplogroup L2.

In brief, haplogroup L2 subdivides in 5 main subhaplogroups (Figure 11): L2a, L2b, L2c, L2d and L2e. Subhaplogroup L2a is by far the most common and widespread in Africa (62 % of total L2) [129], with a frequency over 40 % in Tuareg from Nigeria and Mali [32, 130], Northern

Cameroon [131], Western Pygmies from Gabon [132] and Bantu groups from Mozambique [133]. Due to its wide distribution, its geographic origin is difficult to determine [134].

L2b-d subhaplogroups are believed to be restricted to Western and West-Central Africa [131, 134] and seem to be absent in East Africa and very rare in the South [130, 133]. L2e is the most recent defined clade of L2 with less than 15 complete sequences available. With so few sequences it is difficult to assess the origin of this branch but, since it is very basal (the first split of L2 separates L2e from L2a-d) it most likely had its origin in the West as well.

The demographic history of L2 is not yet completely understood, especially concerning the age of the expansion into Eastern Africa, a region that might have acted as a refuge during some severe episodes of climate oscillations over the last hundred thousand years [135]. It is believed that the expansion of L2 to the East, similarly to the one to the South, was related to movements of Bantu-speaking populations. However, in the region of East Africa with the highest frequency of L2 (over 30 %, in the area of Sudan and Ethiopia) [5] there are no records of Bantu groups. Furthermore, recent evidence from HVRI [5] suggests that this haplogroup might have first expanded to Eastern Africa much earlier, possibly due to the improvement of climate conditions during the early Holocene. This signal was also observed with Bayesian analysis of L2 (and L2a) complete sequences [136].

Particular clades of L2a and L2c suggest an expansion, possibly along the Sahel corridor, after the LGM [32]. The same pattern is observed in branches of other haplogroups such as L0a, L1b and L3f [4, 5, 32].

Despite being spread across different regions, most of the haplogroup L2 sequences available in online databases are either from West-Africans or from African-Americans, which may have biased some studies in the past.

1.4.3.1. Haplogroup L2 over the world

Although typical of Africa, L2 has dispersed recently to other continents and is particularly frequent in America. During the transatlantic slave trade millions of Africans from very distinct regions were forced to move to the American continent. Since L2 is spread across Africa at high frequencies, especially in the West and West central Africa (where most of the black slaves were from [26]), it is also very common today amongst African-Americans (~19 %) [27] (Figure 12).

In Europe, despite some geographic variation, L haplogroups represent less than 1 % of maternal variation and seem to have arrived in the continent in the Post-Glacial period, much before the beginning of the slave trade [137], being L1b the most common [27].

Regarding L2, some branches seem to be typically European. L2a1k seems to be exclusively present in Europe [138] and L2a1l2a haplotypes have been reported amongst Ashkenazi Jews in Europe, namely from Poland, Romania and Russia [139, 140].

In the Iberian Peninsula several studies [141–145] have consistently found L2 lineages, although generally in low frequencies.

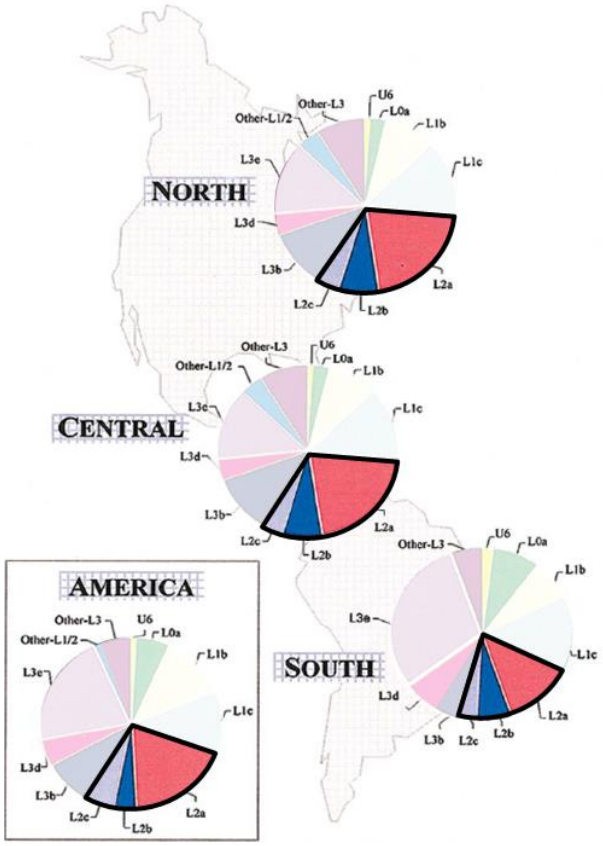

Figure 12. Frequency of haplogroups of recent African ancestry in America (frequency of L2 subhaplogourps highlighted), adapted from [27].

1.4.4. Phylogeographic approach

A phylogeographic approach provides insights on population migrations, however, it is important to bear in mind that the age estimates for lineages involved in migrations are not the exact ages of the migrations. Instead, they indicate when all the mutations that define those lineages arose [146] and serve as upper limits for the expansions.

Moreover, the study of a given lineage does not correspond to the study of a population. A lineage-based approach allows, firstly, to organize the phylogenetic structure of a given haplogroup, so as to decrease nomenclature problems and serve as a tool for quality assessment [126, 147]. But it can also provide insights into the demography of populations, revealing patterns that would otherwise be dismissed by population genetics' approaches.

Examining specific haplogroups has been proved useful before, for instance, to resolve the old debate regarding the Bantu expansion [73].

Although preferred by some authors [102, 146, 148], population-based analysis do not always result in good resolution. Even though we are very likely at the edge of a new era on population genetic (and genomics), the truth is that population studies with complete mtDNA sequences are extremely rare and very recent. Regarding the African continent, apart from the 1000 Genomes project (mostly focused on WA plus one population from Kenya), only populations from SA [57, 75, 149, 150], Burkina Faso [151] and Pygmy groups from CA [152] were extensively sampled for complete sequencing.

Most of the population-based studies rely instead on HVRI (or, at best on CR), which due to its high mutation rate accumulates variations very rapidly [102, 122]. But also due to this rapid rate, homoplasy and recurrence are very frequent [90], resulting in distant lineages sharing similar HVRI diagnostic positions or in the loss of “intermediate” polymorphic states due to back mutations. Therefore, the level of resolution of the majority of population studies based only on HVRI is very low.

In this sense, a phylogeographic approach aimed at specific lineages might help to uncover some interesting patterns in populations, as shown earlier for the Bantu expansion. In this case, a lineage approach allowed to identify some mtDNA (L0a, L2b, L3b and L3e) and Y-chromosome (E1b1a, E1b1a7 and B2b) haplogroups among Bantu groups [32, 76–78] which suggested that the expansion of Bantu languages was due to the migration of some populations, rather than just a cultural diffusion [74].

Nevertheless, when looking at a specific mtDNA haplogroup, one is only looking at the history of that haplogroup, a small fraction of all maternal variation, which is, by itself, a small fraction of all the variation contained within a population. Therefore, combing a lineage-based approach with population analysis, allows to confirm the patterns observed in the phylogenetic trees.

2. Aims

The main goal of this study was to reconstruct the phylogeny of mtDNA haplogroup L2 by analysing samples representative of the diversity of this haplogroup in different African regions, without biasing the analysis towards a given geographical area. In this context, we aimed to:

- Focus our sampling mainly in Eastern Africa, a region poorly genetically characterized, by sequencing 25 complete mtDNA genomes, together with additional samples from other regions;

- Increase the resolution of L2 internal phylogeny, in order to infer expansions associated to haplogroup L2, both within (e.g.: post-glacial migrations, Bantu expansion) and out of the African continent (e.g.: the transatlantic slave trade);

- Estimate node ages by applying different phylogenetic methods: Maximum Parsimony, Maximum Likelihood and Bayesian inference;

- Infer population patterns of Bantu groups, particularly in the Great Lakes region, by performing complementary analysis of haplogroup L0a and a population-based approach.

A resolved phylogeny might be a useful tool for future studies in various fields of research, such as evolutionary, clinical and forensic genetics. Increasing the resolution of the internal structure of L2 will also allow minimizing nomenclature problems in the future.

3. Methods

3.1. Sampling

Initially, 51 samples representative of the diversity of haplogroup L2 (confirmed by a diagnostic mutation, *see 3.2.2*) in different African regions were targeted for complete mtDNA sequencing (Figure 13).

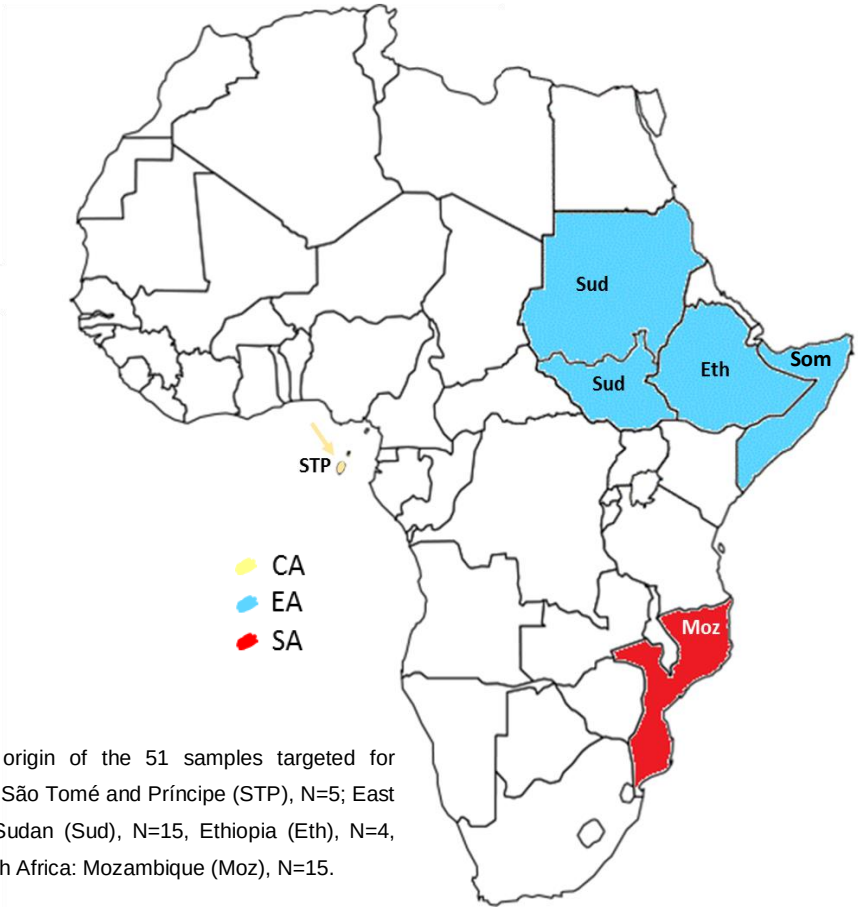

Figure 13. Geographical origin of the 51 samples targeted for sequencing. Central Africa: São Tomé and Príncipe (STP), N=5; East Africa: Sudan and South Sudan (Sud), N=15, Ethiopia (Eth), N=4, Somalia (Som), N=12; South Africa: Mozambique (Moz), N=15.

There was an urge in enriching sampling in EA, therefore, most of the samples selected in this study (31 individuals) were from EA: four individuals from Ethiopia, 12 from Somalia and 15 from Sudan and South Sudan (here considered as one unique population, since the samples were collected before the independence of South Sudan). By the beginning of this project there

were very few L2 complete sequences available from SA, being the majority of them from a recent study on Zambian population [150] and a few occasional samples from Mozambique. Since sequences from the South would be important to disentangle patterns related to the Bantu expansion, we selected 15 samples from Mozambique, a region where L2 seems to be particularly frequent [5]. However, two other studies with complete sequences from SA were published during the course of this work [57, 149], despite not including any sequences from Mozambique. Additionally, five samples from São Tomé and Príncipe (STP), an important stepping-stone for the transatlantic slave trade [16], located in CA, were also selected for sequencing.

From these initial 51 samples targeted, seven (one from Mozambique and six from Sudan) were later excluded, either due to difficulties in amplification and/or sequencing, or due to quality issues (e.g.: ambiguities that could not be resolved and/or many missing haplogroup-defining polymorphisms).

All the samples belonged to unrelated individuals who gave informed consent. Sudanese and Ethiopian samples were collected from emigrants in Dubai and Somali samples were from refugees in Yemen. Samples from Mozambique and STP were collected from the general population. Unfortunately, no information regarding ethnic and/or linguistic affiliation of the individuals is known.

These sequences were combined with more than 750 complete L2 mtDNA sequences (mostly from West-Africans and African-Americans) available on online databases, namely GenBank® and 1000 Genomes. For comparative purposes, analysis of L0a phylogeny (based on published whole-mtDNA sequences) were also performed, since phylogeographic patterns associated to Bantu groups seem to be similar in some branches of both haplogroups [4].

Table 1. Colour scheme and codes for the geographic regions considered.

Region	Code	Colour scheme
America	AM	**America (AM)**
West Africa	WA	**West Africa (WA)**
Central Africa	CA	**Central Africa (CA)**
Southern Africa	SA	**Southern Africa (SA)**
North Africa	NA	**North Africa (NA)**
East Africa	EA	**East Africa (EA)**
Arabian Peninsula/Near East	AP/NE	**Arabian P./Near East (AP/NE)**
Asia	AS	**Asia (AS)**
Europe	EUR	**Europe (EUR)**
Unknown	unknown	**unknown**

The dataset was divided in eight geographic regions (with a more detailed division for Africa), plus an extra category for unknown geographic origin (Table 1).

3.2. Genotyping strategy

Since the molecule of human mtDNA has a length of ~16,569 bp, amplification and sequencing reactions were performed in several steps. The primers (Annex 1) were designed to divide the molecule in 32 overlapping fragments of ~600 bp [153] and were used both for amplification and sequencing reactions.

3.2.1. Pre-PCR protocol

DNA was already extracted for previous studies by Phenol-Chloroform [154] or Chelex®100 (Bio-Rad) [155] methods and the DNA samples were stored at Ipatimup at -20 ºC.

In order to guarantee the yield of further amplifications, since the quantity and quality of the extracted DNA were very low (especially in those samples that have been long-term stored in contact with Chelex®100), a whole-genome amplification reaction with illustra Ready-To-Go GenomiPhi HY DNA Amplification Kit (GE Healthcare) was first performed, according to manufacturer's instructions.

Due to viscosity, these amplification products were diluted in Milli-Q water and 7 samples were tested with a standard PCR (Figure 14) at two different dilutions (1:4 and 1:10). Dilution 1:10 was chosen for further reactions since the amplification was successful for all the samples and, due to its lower viscosity, was easier to handle.

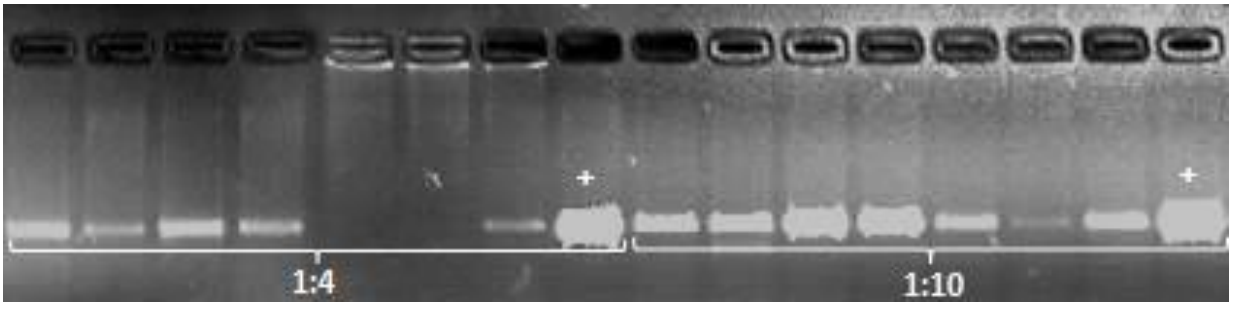

Figure 14. Confirmation of whole-genome amplification by PCR, at two different dilutions (1:4 and 1:10). At 1:4, two samples were not amplified, whereas at 1:10 the amplification was successful for all the samples tested. The positive controls (+) were amplified at both dilutions.

3.2.2. Haplogroup confirmation

Initially, the fragment 18, which encompasses a diagnostic mutation for haplogroup L2 (position 9221), was amplified and sequenced, in order to confirm the previous classification based on HVRI.

In one sample (SUD43) position 9221 did not present the expected polymorphism due to a back mutation. In this case, a different fragment encompassing position 8206 (also diagnostic for L2) was sequenced and the haplogroup was confirmed.

3.2.3. Amplification

PCR conditions were optimized, according to the specifications of the *Taq* DNA Polymerase used (final optimized conditions on Table 2). PCR reaction consisted of 2.5 µL of 10x Reaction Buffer, 1.7 µL of 50 mM $MgCl_2$, 0.7 µL of 10 mM dNTPs NZYMix, 2.5 µL of 2.5 mM primer forward, 2.5 µL of 2.5 mM primer reverse, 0.2 µL of 5 U/µL NZYTaq DNA polymerase and Milli-Q water up to a final volume of 24 µL, to which 1 µL of template DNA (diluted 1:10) was added.

In some samples, primers for fragments 5, 9 and 14 amplified numts, instead of mtDNA. In order to overcome this problem, a different strategy was adopted, which consisted on two PCR reactions.

Table 2. Program conditions of general PCR reactions.

Cycle step	Temperature (ºC)	Time	No. of cycles
Initial extension	94	2’	1
Denaturation	94	30’’	
Primer annealing	57	1’	35
Extension	72	30’’	
Final extension	72	10’	1

Firstly, a PCR with a different combination of primers (Table 3) and a lower annealing temperature (Ta=55 ºC) was performed. With such a low Ta nonspecific bands on the gel were fairly common. Purification was performed as described ahead and the purified products were

used as the template for the second PCR. The PCR mix for the second amplification contained the specific primers for the targeted fragment (5, 9 or 14) and was performed at a higher Ta (Ta=59 ºC), to ensure specificity of the PCR product and guarantee that only mtDNA was amplified. The final products were ~600 bp, correspondent to the expected size (with no extra nonspecific bands), and were confirmed (by alignment to rCRS) to be the targeted region of the mtDNA molecule.

Table 3. Primer combinations used in the first PCR to solve the amplification of numts. Primers named as in Annex 1 (primer sequences listed in Annex 1).

Target (Fragment)	Primer combination (PCR 1)	Primer combination (PCR 2)
5	P4F + P5R P5F+ P6R	P5F + P5R
9	P8F + P9R P9F + P10R	P9F + P9R
14	P13F + P14R P14F + P15R	P14F + P14R

Amplification was always confirmed by 1.5 % agarose gel electrophoresis (Agarose Electrophoresis Grade stained with GreenSafe Premium, both from NZYTech), visualized under UV light, with ChemiDoc™ XRS+ System (Bio-Rad). DNA purification was executed according to illustra MicroSpin S-300 HR Columns protocol (GE Healthcare).

3.2.4. Sequencing

Sanger sequencing was performed with Big Dye® Terminator v.3.1 Cycles Sequencing Kit (Applied Biosystems), accordingly to the conditions in Table 4. The sequencing reaction (final volume = 5 µL) consisted of 1 µL of BigDye® Terminator v3.1 Ready Reaction Mix, 1 µL of 5X Sequencing Buffer, 0.5 µL of 2.5 mM Primer (forward or reverse) and 2.5 µL of purified DNA. The H-strand was sequenced with the reverse primer whenever necessary to resolve ambiguous positions or length heteroplasmy.

After column purification with Sephadex G-50 M DNA Grade (GE Healthcare) to clean the excess of labelled ddNTPs, and DNA denaturation with Hi-Di™ Formamide (Applied Biosystems®), samples were delivered to IPATIMUP Sequencing Service, responsible for handling the 3130XL Genetic Analyzer sequencer (Applied Biosystems®).

Table 4. Sanger sequencing program conditions.

Cycle step	Temperature (ºC)	Time	No. of cycles
Initial extension	96	4’	1
Denaturation	96	15’’	35
Primer annealing	50	9‘’	
Extension	60	2’	
Final extension	60	10’	1

3.3. Haplogroup affiliation

Sequences were compared to rCRS [95] using Geneious v.5.4 [156] and polymorphisms were manually checked and annotated according to the nomenclature in PhyloTree (Build 16, February 2014) [122]. HaploGrep [157] was used to assign each sample to an haplogroup.

3.4. Phylogenetic reconstruction

Firstly, the databases of complete L2 and L0a mtDNA sequences were updated. A total of 335 newly released sequences both from the 1000 Genomes project [24] – 140 samples from phase 2 and 3 (March 2012 and June 2014, respectively) – and from recent publications [57, 139, 149, 150, 158, 159] were added to L2 database. Regarding L0a, an updated L0 tree was recently published [4], therefore only 128 new sequences [57, 149] were added.

Together with the 44 samples from this study, a total of 801 L2 complete sequences were analyzed (Annex 2). For haplogroup L0a, the analysis included a final dataset of 303 samples (Annex 3). MtDNAGenSyn software [160] was always used to convert sequences into haplotypes.

Phylogeny was reconstructed based on preliminary reduced-median network analysis [161] with Network v.4.611 (www.fluxus-engineering.com), which suggested a branching order that was manually confirmed more parsimoniously, according to the structure in PhyloTree (Build 16) and considering the frequency of each mutation, as reported in [37]. Insertions at

positions 309 and 315, indels (insertions/deletions) between positions 515 and 522 and hotspots at 16182, 16183 and 16519 were not considered for any of the analysis performed. The final trees were manually drawn in an Excel file (Microsoft® Office 2013) and in *.xml* format.

3.4.1. Age estimates

In order to estimate the TMRCA of L2 and L0a internal clades, both maximum parsimony (MP), using *rho* (*ρ*) statistics, and maximum likelihood (ML) analyses were performed. In addition to the positions removed for phylogenetic reconstruction, all indels were also excluded for the following analysis, since this type of variation is not considered by the models used for age calculations.

ρ statistics [162] estimates the average of mutational steps from a given ancestral node to the tips of the phylogeny, purely based on a given mutation rate, and not including any evolutionary model. The mutation rate applied considered one substitution in every 3,624 years, correcting for purifying selection and the synonymous mutation rate was of one substitution in every 7,884 years [37]. Standard errors were estimated as in [163].

ML estimates of branch lengths were performed using baseml program included in PAML (Phylogenetic Analysis using Maximum Likelihood) package [164]. In opposition to *ρ* statistics, which only takes into account the structure of the phylogeny, PAML considers both the structure of the tree, as well as the entire dataset of sequences, in order to apply an evolutionary model. We assumed the HKY85 (Hasegawa, Kishino and Yano, 1995) mutation model [165] with gamma-distributed rates (discrete distribution of 32 categories). HKY85 distinguishes transitions from transversions and allows different nucleotide frequencies, having been previously successfully tested for large mtDNA datasets [37]. Two partitions were considered so as to differentiate the fast evolving HVRI and HVRII from the rest of the molecule.

Since evidence of violation of the molecular clock was previously reported for African haplogroups, including for L2 [126, 129, 166], we conducted the PAML analysis for this haplogroup both with and without a molecular clock and performed a likelihood ratio test, which indicated deviations to the molecular clock.

In addition, we estimated ages of L2 internal nodes using BEAST v.1.8.0 [167] (100,000,000 interactions with a burn-in of 10,000,000 steps), applying both a strict and a relaxed molecular clock (which allows rate variation across lineages) and a mutation rate of

2.6186 x 10^{-8} substitutions per site per year (calculated previously for haplogroup L3 [168]). We compared both analysis by calculating a Bayes factor, which showed very strong differences – 2 $\log_e(B10)$ = 13 [169]. The final Bayesian tree, based on a relaxed clock, was created with TreeAnnotator v.1.8.0 (included in the BEAST package) and node ages were visualized with FigTree v.1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/).

3.4.2. Expansion times

Variations in the effective population size (Ne) associated to haplogroup L2 over time were assessed with Bayesian Skyline Plots (BSPs) [170], obtained with BEAST v.1.8.0 and visualized with Tracer v.1.6 [171]. BEAST (Bayesian Evolutionary Analysis by Sampling Trees) uses a MCMC (Markov Chain Monte Carlo) approach to explore the space of genealogies, by randomly sampling from the posterior distribution of a given set of evolutionary parameters (using a Metropolis–Hastings algorithm), according to model parameters and a set of molecular sequences [172].

We ran 100,000,000 interactions, with samples drawn every 10,000,000 MCMC steps (initially burn-in of 10,000,000 steps). HKY85 model (gamma-distributed rates) together with a relaxed molecular clock [173] for a mutation rate of 2.6186 x 10^{-8} substitutions per site per year [168] was employed. A 25-year generation time [174] was assumed and the samples were organized in major monophyletic groups to resemble main subhaplogroups (L2a, L2a1, L2b, L2c, L2d, and L2e), in order to guarantee a tree structure similar to our phylogenetic reconstruction and allow direct comparison among different analysis.

Apart from the BSP for the entire sub-Saharan African dataset of complete L2 sequences, additional regional BSPs (for WA/CA, SA and EA) were computed for higher geographic resolution.

3.5. Frequency distribution maps

In order to visualize their geographic distribution within Africa, frequency distribution maps (based on HVRI data) for major L2 subhaplogroups (L2a, L2b, L2d, L2e and L2*) were

constructed with Surfer® v.8 (Golden Software) using Kriging algorithm. L2c is not distinguishable only by HVRI polymorphisms [32], therefore its frequency was calculated as L2*.

The dataset included 13910 HVRI samples from 39 different populations (Figure 15). This dataset contains published HVRI African sequences, plus the HVRI segment of complete mtDNA sequences from 1000 Genomes and recent population studies [57, 75, 149–151, 175].

A table with the geographic coordinates of the points considered to build the maps is provided in Annex 4. We considered the capital city of the countries sampled as the reference points for the maps, since precise geographic information is not available for most of the populations sampled and some of the samples were not collected in their original countries (such as, for example, the samples from EA sequenced in this study). Pygmy groups were not considered, due to their unusual haplogroup composition that could bias the results.

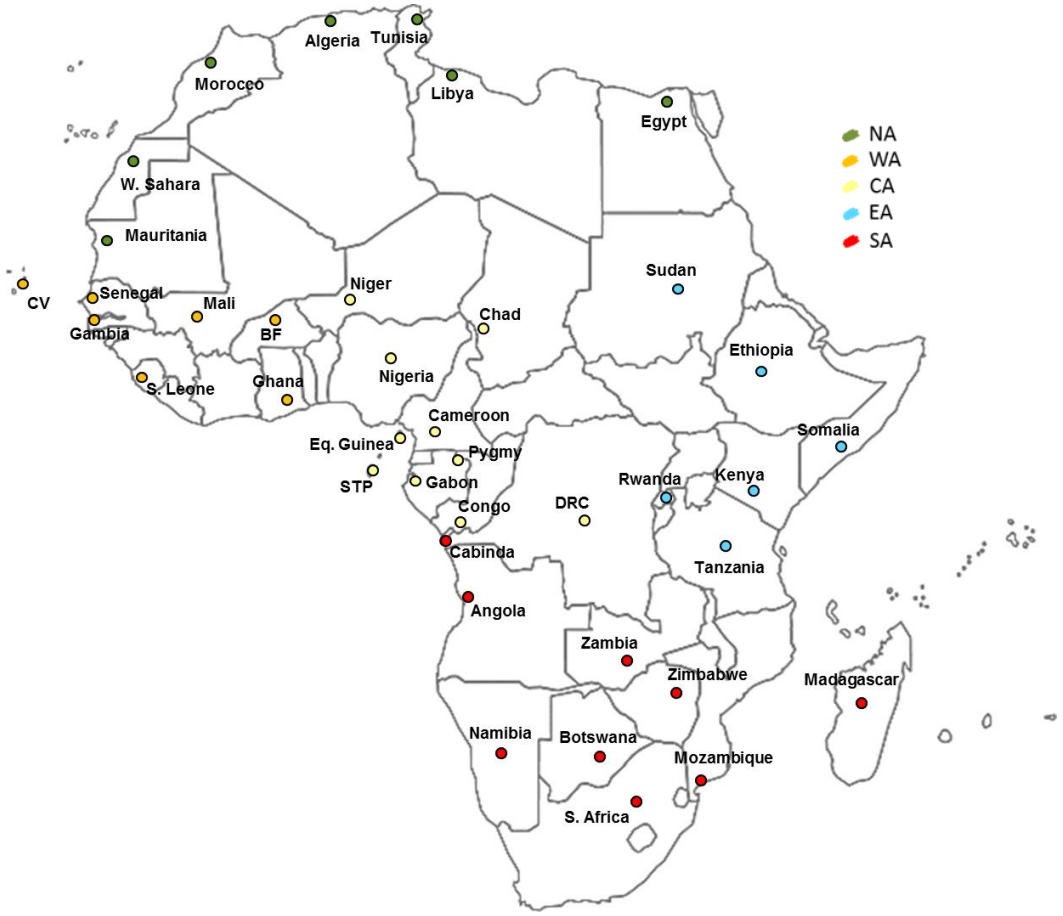

Figure 15. Geographical points considered for the 39 populations included in the dataset used to build haplogroup L2 frequency distribution maps. Western Pygmies (W. Pygmies) and Mbuti Pygmies from Democratic Republic of Congo (DRC) were excluded from the dataset used to build frequency distribution maps, due to their particular haplogroup composition that could potentially bias the results using Kriging algorithm; they were only included in the dataset used to assess haplogroup composition, where no algorithm was assumed. Geographical coordinates and references in Annex 4. Abbreviations: CV – Cape Verde , DRC – Democratic Republic of Congo, Eq. Guinea – Equatorial Guinea, STP – São Tomé and Príncipe, S. Africa – South Africa , S. Leone – Sierra Leone, W. Sahara – Western Sahara.

3.6. Population comparisons

From the dataset used to compute the frequency distribution maps, only populations with more than 30 individuals (N>30) and with ethnic and/or linguistic information available were considered for population-based analysis. However, the general populations that were sampled for the L2 phylogeography analysis (Ethiopia, Mozambique, STP, Somalia and Sudan) were also included. Since the main goal was to infer the relationships between sub-Saharan groups (particularly between Bantu and Eastern groups), populations from NA, Pygmy and Khoisan groups were not included in these analyses.

Genetic distances between pairs of populations (F_{ST}) based on a 276 bp fragment of HVRI were computed with Arlequin v.3.5.1.3 [176] (10,000 permutations). The fragment considered corresponds to the smaller range common to all publications (from position 16090 to 16365). Relationships among populations were represented by multidimensional scaling (MDS) plots (obtained with IBM® SPSS Statistics v.22, with PROXCAL algorithm), based on Slatkin's linearized F_{ST} [177].

Same ethnic groups from different countries and/or different studies, as well as different ethnic groups from the same country were merged and considered as one unique population whenever F_{ST} was not significant in a previous analysis. Bonferroni correction to *p*-values was always applied. The final MDS plot includes 55 populations (Annex 5), corresponding to 4880 individuals. In addition, genetic distances excluding L2 and L0a sequences from the dataset were assessed and a second MDS plot constructed, based on a total of 3323 individuals.

We assessed mtDNA haplogroup composition in West, Central, East and Southern Africa (both by region and by country). Reduced-median network analysis [161] of haplogroups L0a and L2 (based on HVRI) was additionally performed. Since these two haplogroups are both present in Eastern, Central and Southern populations, they are expected to provide insights on the migration patterns of population groups in those regions, so as to distinguish patterns associated to the Bantu expansion from other migrations. Founder ages of L2a main nodes in EA was performed considering the mutation rate previously calculated for HVRI [105], scaled to the fragment considered here (position 16090 to 16365).

4. Results and discussion

4.1. L2 phylogeography

The complete phylogenetic tree for L2 haplogroup is shown in Annex S1 (digital support), including ρ age estimates (considering both the complete genome and the synonymous clocks), ML age estimates for the main nodes and Bayesian age estimates based in a relaxed molecular clock.

Age estimates based on ρ statistics are more prone to bias due to variations in Ne (population expansions, bottlenecks, founder effects) and rate heterogeneity among sites and/or lineages [178]. Therefore, ML and Bayesian analysis, which take an evolutionary model into account and are not dependent on Ne, are expected to provide more reliable results. Bayesian estimates have an additional advantage of allowing rate variation between branches, which is crucial in haplogroup L2 due to clock violations [126, 129, 166].

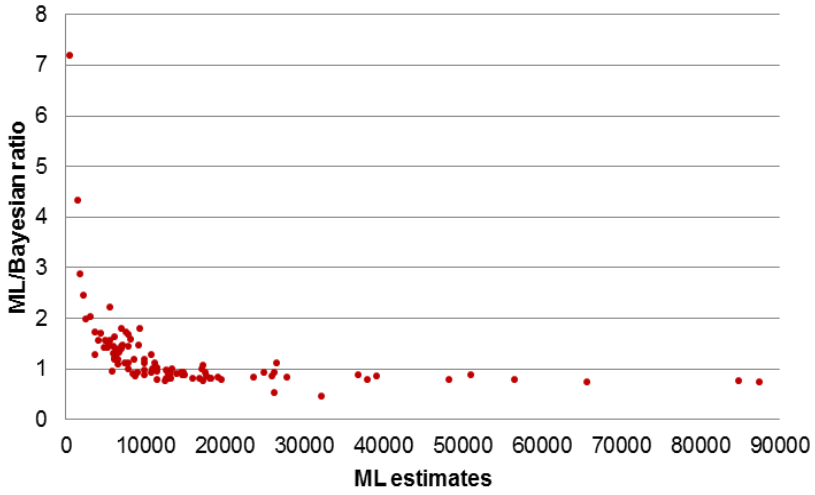

Figure 16. Graph representing variations in the ratio between ML and Bayesian node age estimates compared to the ML age of the branches.

However, for younger branches, there is a risk of overestimating node ages, since it does not account for purifying selection in a time dependent curve (Figure 16). Moreover, in this particular case, we have used as calibration point the TMRCA for haplogroup L3 (ML node age

~70.2 ka), which is much older than the youngest L2 clades and might also result in the overestimation of TMRCAs in the youngest branches.

Therefore, along this chapter, whenever a specific branch is mentioned, the age estimate provided will always be the result of ML and Bayesian analysis (ML; Bayesian estimates). All the TMRCA estimates for the L2 clades mentioned in this chapter are shown in Table 5.

Table 5. Age estimates of L2 clades mentioned in this chapter.

Clade	Bayesian relaxed clock	ML whole mtDNA	*ρ* whole mtDNA	*ρ* synonymous age
L2	78200 [66700 - 92100]	99800 [83600 - 116300]	77700 [57800 - 98300]	86500 [51700 - 121300]
L2a-d	73100 [60400 - 85100]	94100 [78600 - 110000]	70200 [52300 - 88700]	80000 [40300 - 119600]
L2a	65500 [53700 - 78600]	84900 [67500 - 103000]	58900 [38600 - 80200]	80000 [40300 - 12000]
L2a1	29700 [22300 - 36500]	26600 [18000 - 35700]	23300 [17800 - 28900]	24800 [15700 - 34000]
L2a1+143+16189 +16192	14300 [13000 - 17900]	12900 [9100 - 16800]	13000 [8600 - 17600]	124000 [6900 -17900]
L2a1a2a1a	8100 [6400 - 10200]	5400 [3200 - 7500]	7000 [3500 - 10600]	5300 [2300 - 8200]
L2a1a2b	7400 [4400 - 10900]	5200 [1700 - 8900]	5700 [2100 - 9300]	10500 [1600 - 19400]
L2a1a3b	7300 [3500 - 8500]	6600 [2900 - 10500]	10600 [4000 - 17400]	7900 [0 - 16800]
L2a1c2	12270 [9200 - 14700]	12800 [8600 - 17100]	15100 [9600 - 20800]	10300 [5700 - 15000]
L2a1e1	7900 [4900 - 11200]	7800 [3400 - 12300]	8400 [4900 - 12100]	11800 [4400 - 19200]
L2a1f1	10100 [7600 - 12300]	7000 [4700 - 9300]	7500 [5000 - 10100]	11100 [5400 - 16900]
L2a1l1	11500 [9200 - 13000]	11100 [5000 - 17500]	10300 [6700 - 14100]	10600 [4800 - 16500]
L2a1l2a	6500 [4100 - 8900]	1500 [0 - 3800]	1800 [0 - 3900]	600 [0 - 1800]
L2a1l2a1	3600 [500 - 5400]	500 [0 - 3100]	1000 [0 - 2500]	-
L2a1k	8800 [4900 - 13500]	9800 [3300 - 16500]	10600 [3200 - 18400]	7900 [0 - 18800]
L2a1d1	10000 [6400 - 13400]	10700 [5700 - 15800]	13900 [7100 - 21000]	17300 [3900 - 30800]
L2a1d2	13100 [10700 - 16600]	15900 [10000 - 21900]	17400 [6000 - 29400]	10600 [0 - 24200]
L2a2a	12600 [8700 - 16100]	13900 [7600 - 20400]	10600 [5200 - 16200]	17100 [3700 - 30500]
L2a2a1	9000 [5500 - 12500]	11400 [5600 - 17400]	9900 [4500 - 15500]	13800 [2200 - 25400]
L2a2'3'4	32300 [23700 - 40900]	36800 [25400 - 48800]	37000 [25200 - 49400]	43200 [23900 - 62600]
L2a5	45500 [33900 - 57900]	56600 [41600 - 72100]	49900 [33700 - 66900]	57200 [28200 - 86300]
L2a5a	30200 [21200 - 29100]	38000 [26100 - 50400]	34200 [22900 - 46000]	34700 [17600 - 51800]
L2b'c'd	65700 [52400 - 78000]	87400 [71000 - 104000]	80400 [59300 - 102400]	50000 [29100 - 70800]
L2b'c	48900 [37700 - 60600]	65700 [50100 - 81900]	60000 [43900 - 76600]	39500 [22900 - 56000]
L2b	24500 [19600 - 30500]	26200 [20500 - 32100]	29400 [19300 - 40000]	31600 [15400 - 47900]
L2b1a3	12300 [10100 - 14100]	5500 [2900 - 8200]	4800 [2800 - 6700]	4900 [600 - 9200]
L2b1a4	9600 [5900 - 13000]	12500 [7600 - 17500]	14300 [5200 - 2400]	15800 [1200 - 30300]
L2b2a	7700 [4700 - 10800]	6200 [2000 - 10600]	4800 [1800 - 7900]	6800 [500 - 13000]
L2b3+12011+199	8300 [5000 - 12400]	13600 [6400 - 21100]	19000 [10600 - 27700]	13100 [1600 - 24700]
L2c	18500 [14200 - 23000]	17300 [13900 - 20800]	18500 [14300 - 22800]	15900 [11300 - 20600]
L2c1	12100 [9500 - 14100]	13000 [8300 - 17800]	16700 [9600 - 24100]	15200 [2900 - 27600]
L2a1c4a1	4600 [2000 - 6800]	3600 [0 - 7400]	4600 [300 - 8900]	7900 [0 - 17300]

Table 5. *(continued)*

L2c2	13200 [11400 - 14800]	14900 [11400 - 18400]	16900 [10700 - 22600]	17700 [8000 - 27500]
L2c2a1	7800 [6200 - 1011]	6000 [2000 - 10000]	7900 [900 - 15100]	2300 [0 - 4500]
L2c2b1b	5000 [3000 - 8700]	2500 [0 - 5600]	2600 [700 - 4500]	2300 [0 - 5400]
L2c3	13500 [11000 - 16000]	17300 [10200 - 24700]	14200 [9200 - 19400]	12600 [4200 - 21100]
L2c4	11500 [7900 - 14600]	13100 [7500 - 19000]	9300 [3800 - 14900]	17100 [3200 - 31000]
L2c5	10800 [6800 - 14000]	11400 [4800 - 18100]	7400 [2900 - 12100]	9200 [0 - 19200]
L2d	16100 [12700 - 20200]	19100 [13300 - 25100]	17900 [10200 - 25800]	18900 [6200 - 31600]
L2d1	13900 [11700 - 16600]	16900 [11700 - 22200]	16200 [7200 - 25600]	19700 [2000 - 37400]
L2e	33900 [23900 - 43300]	39000 [28600 - 50000]	32600 [22600 - 43100]	44300 [24700 - 64000]
L2e1	15000 [10300 - 22300]	32000 [20600 - 44000]	23100 [12800 - 33800]	38100 [13900 - 62300]
L2e1a	15000 [10300 - 22300]	18200 [9000 - 27700]	16500 [6400 - 27200]	17700 [0 - 37800]

4.1.1. Main L2 clades

In general, L2 has a very complex structure, with high levels of homoplasy and many recurrent polymorphisms defining different branches, both at HVRI (positions 16093, 16129, 16189, 16292 and 16390) and at HVRII (position 143). The phylogeny reconstruction with the published and our new L2 complete sequences allowed to identify new branches and to increase the resolution in many other clades.

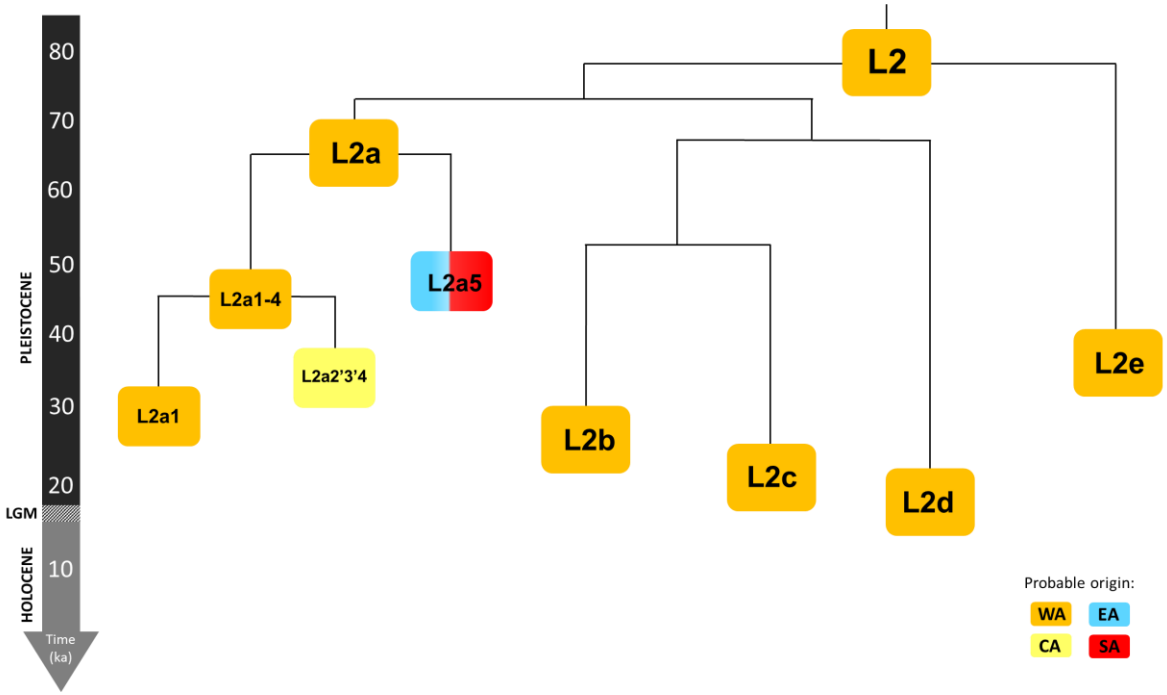

Figure 17. Schematic phylogeny of haplogroup L2 based on Bayesian TMRCA estimates (more accurate for older clades). Colour scheme as in Table 4, corresponding to the probable geographic origin of each clade.

L2 divides into five main branches (L2a-e) (Figure 17), being the split between L2e and L2a-d (L2a'b'c'd) the earliest one (~99.8 ka; 78.2 ka). The subsequent division splits L2a from L2b-d (L2b'c'd) (~94.2 ka; 73.1 ka), followed by the split between L2b'c and L2d (~87.4 kA; 65.7 ka). The latest major split separates L2b and L2c (~65.7 ka; 48.9 ka).

L2a (~84.9 ka; ~65.5 ka), as expected [32, 129], is geographically widespread and highly frequent throughout Africa (Figure 18). It is by far the biggest and most common L2 branch, accounting for ~72 % of total L2 (10 % higher than what was previously described [129]) and with a maximum frequency of ~30 % in Ghana, Sudan and Mozambique (Annex 6). In EA, the least sampled region before this work, L2a is fairly common amongst the complete sequences (~85 % of total L2 found in EA). Specifically regarding our sampling, from the 26 EA samples sequenced, 23 (>88 %) cluster within L2a.

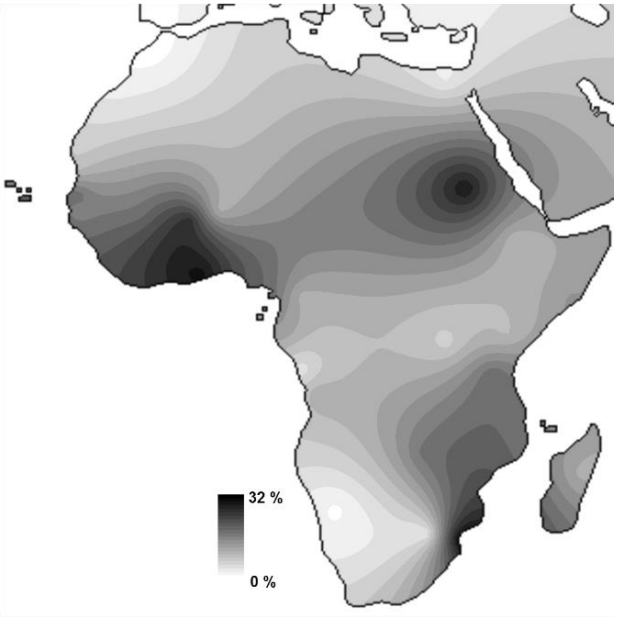

Figure 18. Frequency distribution map for L2a in Africa.

L2a splits into five branches (L2a1-5), being L2a1 (~26.6 ka; ~29.7 ka) the most frequent and most complex one. L2a1 by itself accounts for ~66 % of the entire L2 dataset (corresponding to ~91 % of L2a). In opposition to other L2a clades, which are much more geographically restricted, L2a1 not only harbours samples from all African regions, but has also samples from other continents, including non-African branches, such as L2a1l2a (connected to Ashkenazi Jewish Diaspora [139, 140]), and the exclusively European L2a1k [138] – discussed in detail further ahead. Much of the homoplasy and recurrence observed throughout the L2 tree occurs amongst L2a clades.

L2a2'3'4 (~36.8 ka; ~32.3 ka), on the contrary, seems to have a more Central/Eastern distribution, with a clear association to Pygmy groups. L2a5 *(discussed in detail in 4.1.2.3)*, basal to L2a1 and L2a2'3'4, is practically only present in SA, with no subhaplogroups defined in PhyloTree (Build 16). However the newest published sequences revealed that this is a very ancient (~56.6 ka; ~45.5 ka) and highly structured clade (Figure 27, *see 4.1.2.3*). PhyloTree classification considers nine diagnostic positions for L2a5, but, taking into account these recent sequences, the most parsimonious reconstruction should consider only seven defining positions (being the other two – 3654 and 6497 – diagnostic for a more derived branch).

Regarding L2b (~26.2 ka; ~24.5 ka) – ~10 % of L2 total – it is difficult to unravel any geographic pattern from the phylogeny, with complete sequences from different regions (including from the Arabia Peninsula/Near East and Europe) scattered across different branches. Previous results have reported absence of L2b in EA [44, 129, 130, 134]. Although the majority of complete L2b sequences are either from WA or from African-Americans, samples from EA and NA (in a lower number) are also present within this haplogroup. Some particular clades seem to be connected to specific regions: L2b1a3 (~5.5 ka; ~12.3 ka) and L2b2a (~6.2 ka; 7.7 ka) have a clear southern African distribution, whereas L2b1a4 (~12.5 ka; ~9.6 ka) seems to be non-African. HVRI analysis (Figure 19a) indicates that L2b is distributed essentially in the West, although being present throughout the entire continent at lower frequencies than previously described [134] (maximum frequency in WA ~7 %).

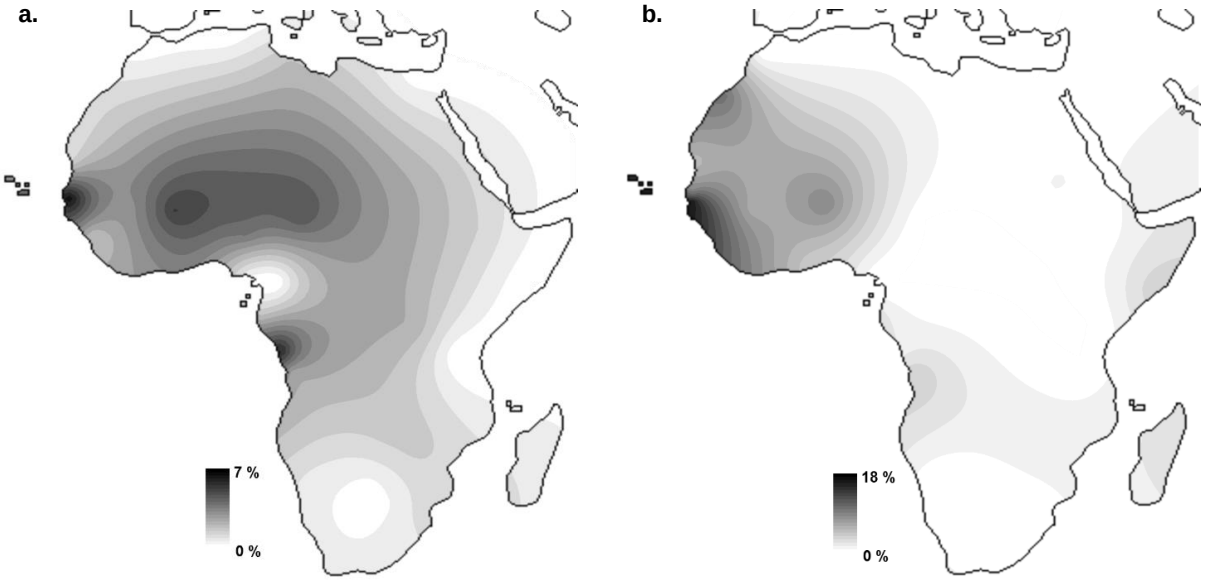

Figure 19. Frequency distribution map for L2b (a) and L2* (b) in Africa.

L2c (~15.3 ka; ~18.5 ka) accounts for ~13 % of our L2 dataset and is essentially western, as reported elsewhere [179, 180], reaching higher frequencies in Gambia, Sierra Leone and Cape Verde (Annex 6). Despite having been described as absent in SA and EA [44, 130, 134], HVRI population data (Figure 19b) shows its presence in these regions, at very low frequencies (<3 %). More importantly, two specific branches – L2c2a1 (-6 ka; ~7.8 ka) and L2c2b1b (~2.5 ka; ~5 ka) – seem to be associated to southern populations and most likely related to the Bantu expansion (discussed ahead in this chapter). However, no complete sequences from EA belong to this haplogroup. A more extensive sampling might allow to discern potential L2c patterns within EA.

L2d (~19.1 ka; ~16.1 ka) and L2e (~39.2 ka; ~33.9 ka) are the smallest and less frequent branches (maximum frequency ~4 %) (Figure 20), representing together less than 5 % of the entire L2 phylogenetic tree. No geographic pattern is discernable within L2d based on the complete sequences presently available. However, in a branch with 20 complete sequences, two are from EA (Sudan and Ethiopia) and one is from Yemen (whose mtDNA gene pool seems to be close to EA [55]). Additionally, HVRI analysis shows that despite being rare (~4 %), L2d is also present in EA. (Figure 20a). Considering that EA is poorly sampled, a more exhaustive sampling in the Horn of Africa would probably allow to increase the resolution of L2d and perhaps allow to distinguish a geographic pattern within this clade.

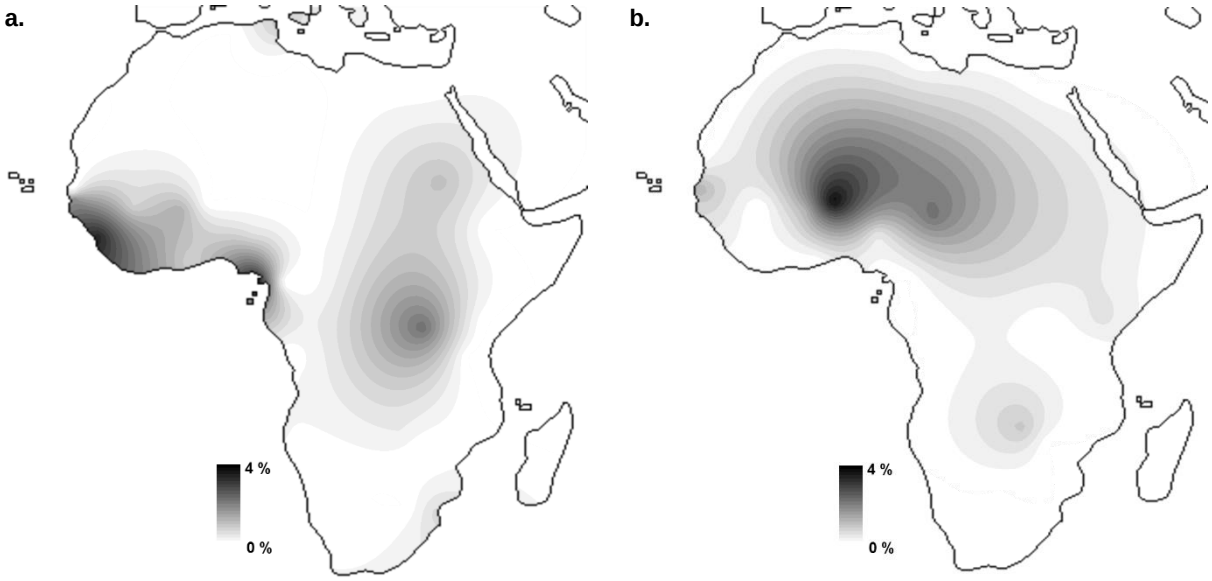

Figure 20. Frequency distribution map for L2d (a) and L2e (b) in Africa.

L2e split from L2a-d corresponds to the earliest division in L2 phylogeny. Although HVRI data display a northern distribution, with the highest frequencies in the Sahel belt (Figure 20b), no complete L2e sequences from this region have been reported. Once again, our analysis revealed higher structure than previously described in PhyloTree, not only within L2e, but also within L2e1 (~32.2 ka; ~23.9 ka). Furthermore, a connection to the Bantu expansion might be patent at L2e1a (which will be discussed further ahead).

4.1.2. Geographic patterns

Across the entire L2 tree, signals of various migrations at different times throughout sub-Saharan Africa, as well as, at a smaller extent, to NA are discernible. Also, the L2 tree clearly

shows movements from Africa towards other continents, being the signal associated to the transatlantic slave trade the most evident.

4.1.2.1. Western and Central Africa

Our results support a western African origin for L2, as previously proposed [5]. All major clades (L2a-e) are present in WA, meaning that L2 was structured very early (before 45 ka), with these main branches appearing in WA during the Pleistocene, before the first L2 migrations. Additionally, all major branches show peaks of frequency in WA (Figures 18, 19 and 20).

Moreover, many big clades (such as L2a1+16189+16192, L2c2, L2d1, L2e1) seem to derive directly from western branches (with occasional American samples). It is not uncommon for western branches to include African-American sequences, since WA was the most common source of African slaves during the transatlantic slave trade [16, 27].

On the other hand, many smaller and more derived branches, such L2a1l1, L2a1c2, L2c1, L2c3, L2c4 and L2c5, are essentially western (again, with sporadic samples from America). These branches might not have been involved in any migrations (or did not leave fertile descendants outside WA), having remained geographically restricted to WA.

Central African sequences are present in all main branches and normally associated to western samples, as expected considering their geographical proximity.

4.1.2.2. Eastern Africa

From the 801 complete L2 sequences contemplated in our analysis, only 39 are from EA, evidencing how poorly represented this region is among the majority of studies. Considering that four samples are from Kenyan Bantu individuals (Luhya), only 35 are truly Eastern African autochthonous lineages. From these, more than half (25 samples) were sequenced in the present study. Apart from one Jewish and the four Bantu samples, no ethnic or linguistic information is known for any other eastern samples, resulting in the loss of potential information regarding Cushitic, Sudanic and Nilotic dispersals.

The majority of eastern African samples cluster with western or central populations, such as in L2a2a (~13.9 ka; ~12.6 ka) and L2a2a1 (~11.4 ka; ~9 ka) – samples from Sudan group together with Chad and Pygmy groups (Figure 21). These observations support a western origin of L2 and a later expansion across the continent, possibly through the Sahel corridor, towards EA.

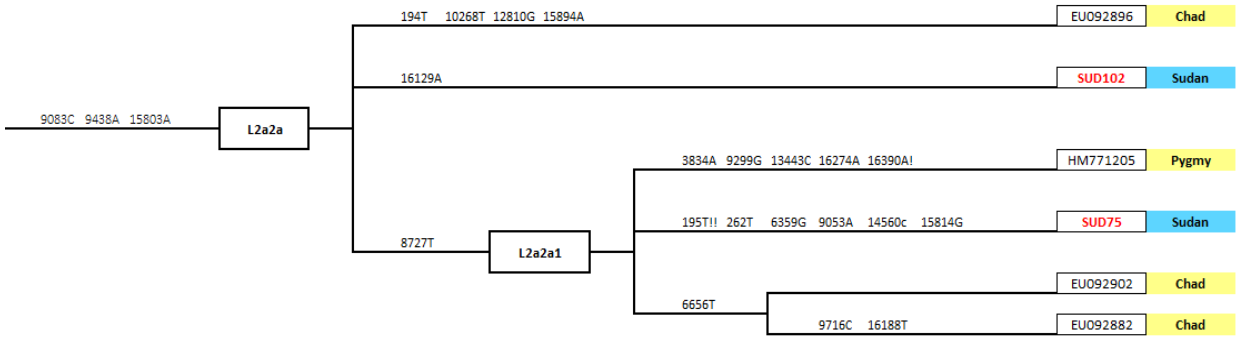

Figure 21. Phylogenetic tree of subhaplogroup L2a2. Polymorphic positions annotated along the branches (mutations in relation to an ancestral state): uppercase for transitions and lowercase for transversions, back mutations indicated with an exclamation mark (two exclamation marks for double back mutations). Samples in red (SUD75 and SUD102) sequenced in this study. More detailed information on these samples on Annex 2.

More importantly, these branches, either east African or east African shared with Western/Central Africa (L2a2, L2a1d1, L2a1+143+16189, L2a1+143+16189+16192 and the unnamed branch with 6743 and 15924, here labelled as L2a1r) all date between 15-8 ka and present EA basal sequences, suggesting an early migration after the establishment of the clades. These results are in agreement with our hypothesis regarding the expansion of haplogroup L2, mostly L2a, towards the East during the early Holocene, adding up to a series of previous evidence [5, 136].

Additionally, proximity between EA and northern African populations seems to have left its imprint in the phylogeny of haplogroup L2. One example is L2a1d1 (~10.7 ka; ~10 ka) (Figure 22), which seems to reflect the historic contact between eastern Africa groups and Egypt [35].

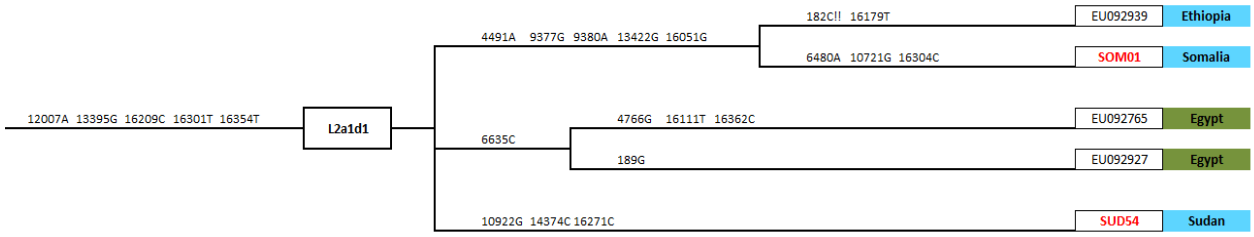

Figure 22. Phylogenetic tree of a subset of subhaplogroup L2a1d1. Polymorphic positions annotated along the branches (mutations in relation to an ancestral state), double back mutations indicated with two exclamation marks. Samples in red (SOM01 and SUD54) sequenced in this study. More detailed information on these samples on Annex 2.

There are also some samples from the Arabian Peninsula and the Levant in the L2 phylogeny that cluster with eastern and northern African samples, such as in L2a1+143+16189+16192 (~12.9 ka; ~14.3 ka) (Figure 23), supporting previous genetic and historical evidence of gene flow in the area [47, 48, 50].

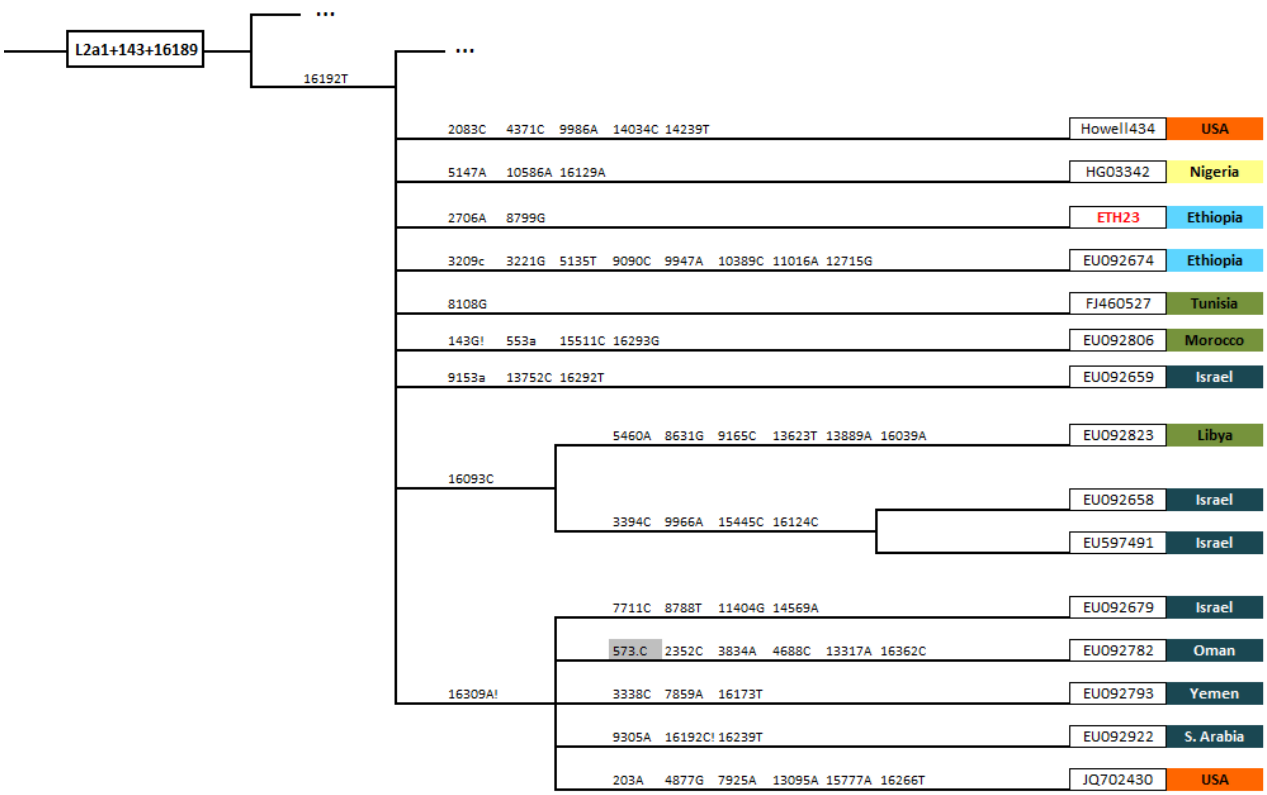

Figure 23. Phylogenetic tree of a subset of subhaplogroup L2a1+143+16189+16192. Polymorphic positions annotated along the branches (mutations in relation to an ancestral state): uppercase for transitions and lowercase for transversions, back mutations indicated with an exclamation mark, insertions indicated by a dot (shaded in grey; not considered for age estimations). More detailed information on these samples on Annex 2.

4.1.2.3. Southern Africa

The newly published complete sequences from SA [57, 149, 150] provided much more detail to the structure of some branches probably associated to Bantu movements. Many examples can be found across our tree falling within the timeframe for the Bantu expansion (<5 ka) [32]. For these recent branches, ML estimates, which account for purifying selection, are most likely more precise in comparison to the overestimated Bayesian node ages.

One of the best examples is clade L2e1a (Figure 24), in which an exclusively southern branch, dating ~1.8 ka (Bayesian estimate: ~5.2 ka), directly derives from a basal sample from Nigeria, probably a reminiscent from the very earliest steps of the Bantu migration, which are thought to have occurred out of the Grassfields region of southeast Nigeria and western Cameroon around this time [79].

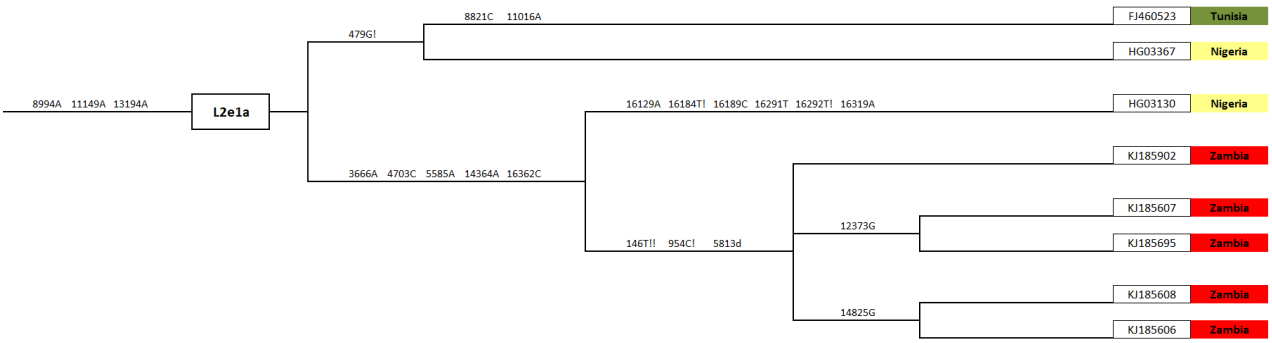

Figure 24. Phylogenetic tree of subhaplogroup L2e1a. Polymorphic positions annotated along the branches (mutations in relation to an ancestral state): two exclamation marks for double back mutations, deletions indicated by the letter "d" (shaded in grey; not considered for age estimations). More detailed information on these samples on Annex 2.

Other examples, as L2c2b1b (~2.5 ka; ~5 ka) (Figure 25) and L2a1d2+5442+16129+7741 (here labelled L2a1d2a) (~3.6 ka; ~7.2 ka) (Figure 26), also deriving from west-central samples, show a clear star-like structure in southern branches, typical from recent and rapid expansions. In the latter, although complete CA samples are missing, HVRI data suggests the presence of the clade in CA (a network is displayed later on this chapter).

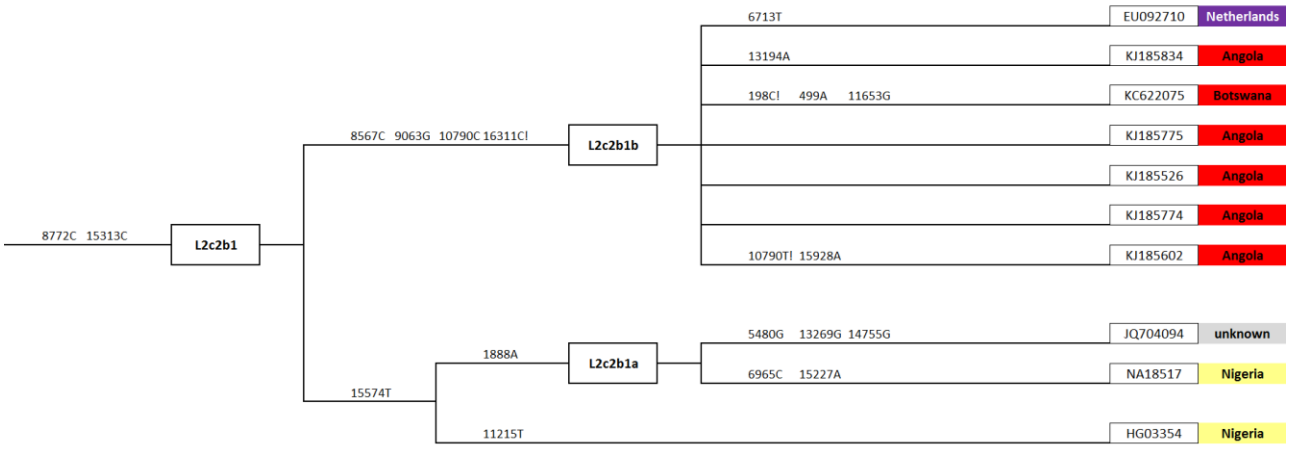

Figure 25. Phylogenetic tree of subhaplogroup L2c2b1. Polymorphic positions annotated along the branches (mutations in relation to an ancestral state), back mutations indicated with an exclamation mark. More detailed information on these samples on Annex 2.

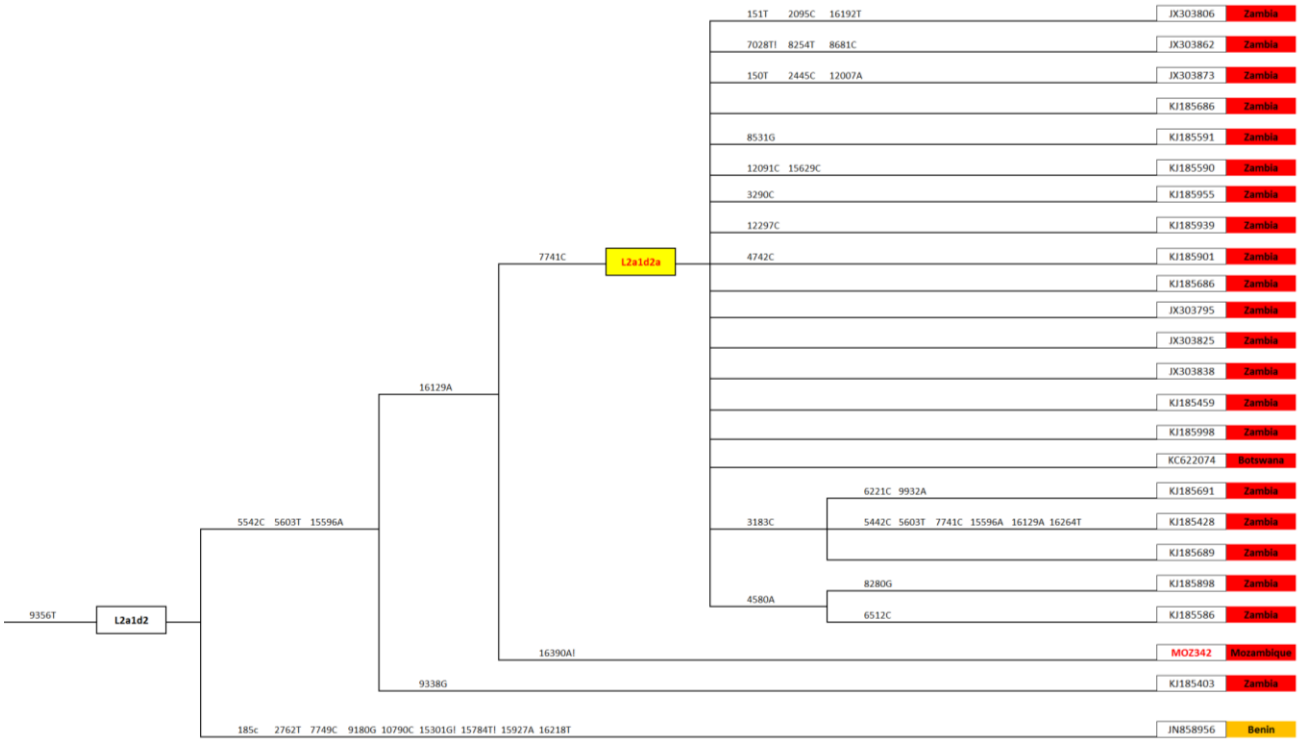

Figure 26. Phylogenetic tree of subhaplogroup L2a1d2, label suggested for a new subhaplogroup (L2a1d2a) shaded in yellow. Polymorphic positions annotated along the branches (mutations in relation to an ancestral state): uppercase for transitions and lowercase for transversions, back mutations indicated with an exclamation mark. Sample in red (MOZ342) sequenced in this study. More detailed information on these samples on Annex 2.

L2a5 (Figure 27), apart from one sample from Spain and another from Bermuda, is typically southern African. However, not only does it not seem to be related to the Bantu expansion, but it is also very ancient (~56.6 ka; ~45.5 ka). L2a5 haplotypes were also found in the HVRI dataset in samples from Sudan, indicating that this haplogroup is also present in EA, even though there is no complete eastern L2a5 sequences included in our analysis. The age of the clade suggests a much earlier expansion of this clade to EA and/or to SA. Movements from Eastern to Central Africa might have carried ancestral lineages of L3e and L3b'd around 50-40 ka [168]. Those movements were probably associated to climate change [181, 182] and it is not unlikely that movements occurred in both directions.

This clade being incorporated by Bantu-speakers represents the major input of autochthonous lineages (either in the East or in the South) apart from the presence of L0d and L0k further south. Another possibility is that this lineage was already present in the early Bantu populations moving South after the standing point in the Great Lakes in Uganda. Unfortunately, mtDNA genetic variation of Uganda is still a blank.

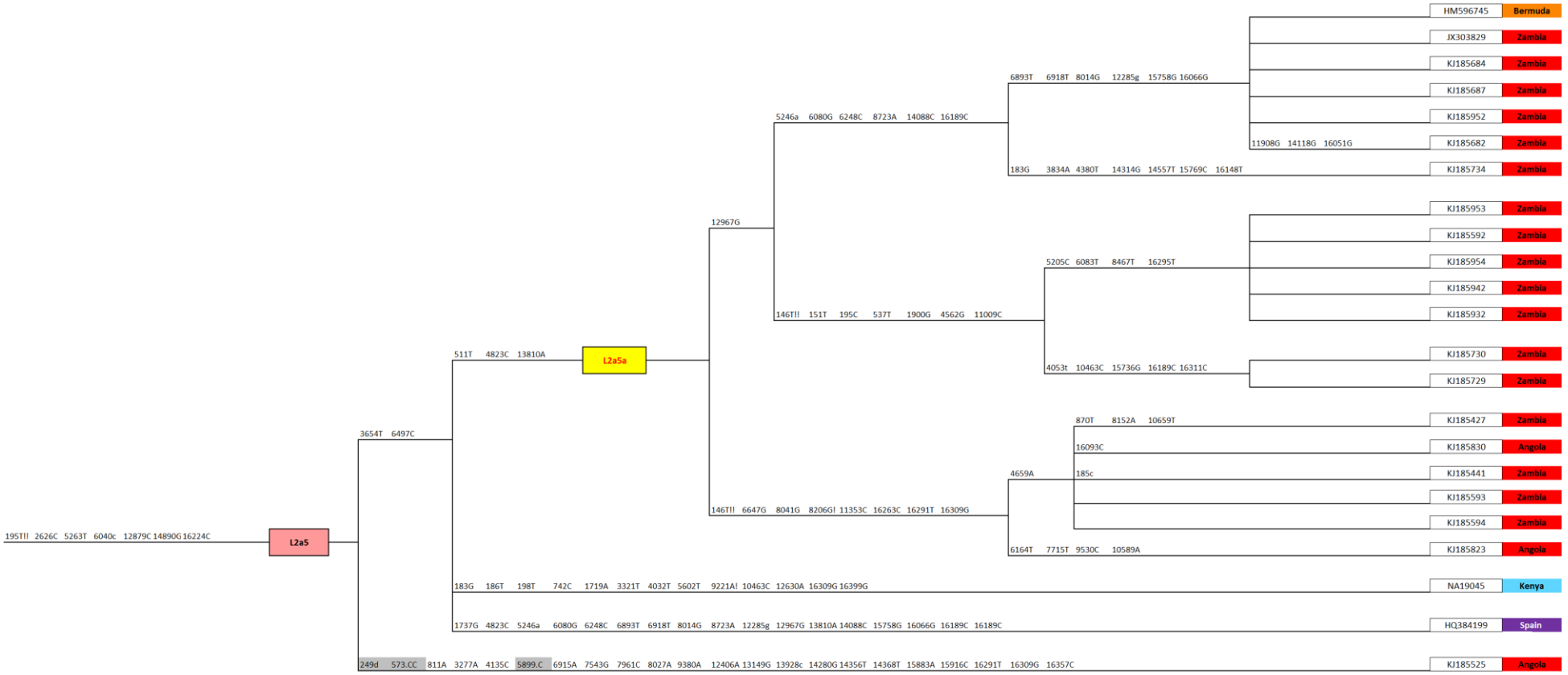

Figure 27. Phylogenetic tree of subhaplogroup L2a5 (with major alterations in comparison to PhyloTree Build 16), label suggested for a new subhaplogroup (L2a5a) shaded in yellow. Polymorphic positions annotated along the branches (mutations in relation to an ancestral state): uppercase for transitions and lowercase for transversions, back mutations indicated with an exclamation mark (two exclamation marks for double back mutations), insertions indicated by a dot and deletions by the letter "d" (both shaded in grey; not considered for age estimations). More detailed information on these samples on Annex 2.

4.1.2.4. Transatlantic slave trade

Between the 15th and 19th centuries millions of people were forced out of Africa due to what came to be known as the transatlantic slave trade. The vast majority was taken to the American continent, to work in European colonies. The slave commerce reached its peak in the Bight of Benin, in WA, known at that time as the Slave Coast [16].

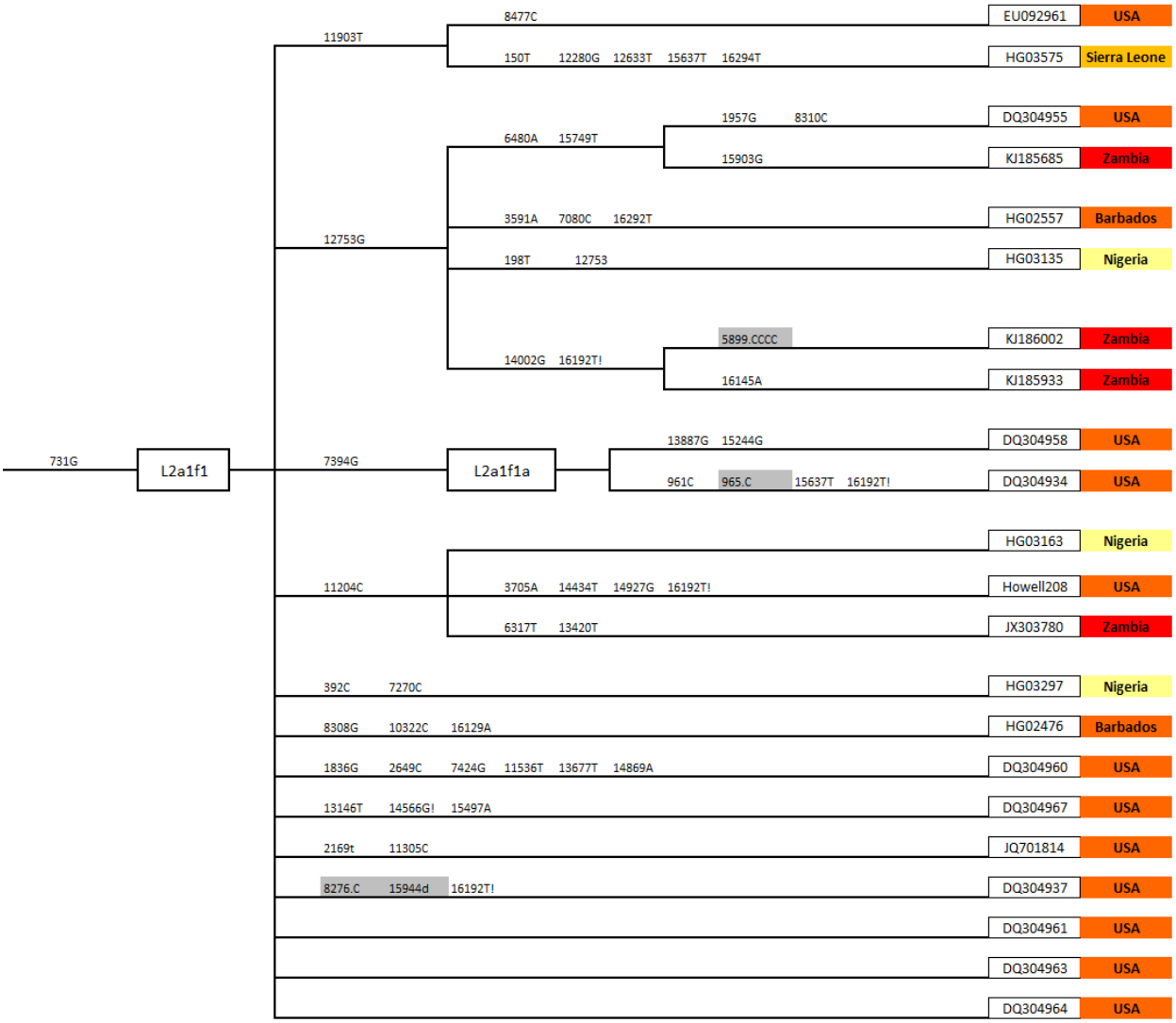

Figure 28. Phylogenetic tree of subhaplogroup L2a1f1. Polymorphic positions annotated along the branches (mutations in relation to an ancestral state): uppercase for transitions and lowercase for transversions, back mutations indicated with an exclamation mark, insertions indicated by a dot and deletions by the letter "d" (both shaded in grey; not considered for age estimations). More detailed information on these samples on Annex 2.

Being so frequent in WA, L2 is also very common amongst African-Americans [27]. Signs of this forced migration can be observed all over the phylogeny of haplogroup L2, particularly in L2a1. Most of the American samples in our tree cluster, as expected [27, 183],

either with WA (e.g.: L2a1a2b, L2a1a3b, L2a1e1 and L2a1l1) or with CA (more precisely with Nigeria), in agreement with historical records. Southern Nigeria was the most populous zone of Africa and is conveniently situated near the Bight of Benin, making it a very suitable source of slaves [16]. L2a1f1 (~7 ka; ~10.1 ka) portraits very well this situation, harbouring a large number of American samples (from the USA and Barbados), together with old samples from Nigeria and WA (Figure 28).

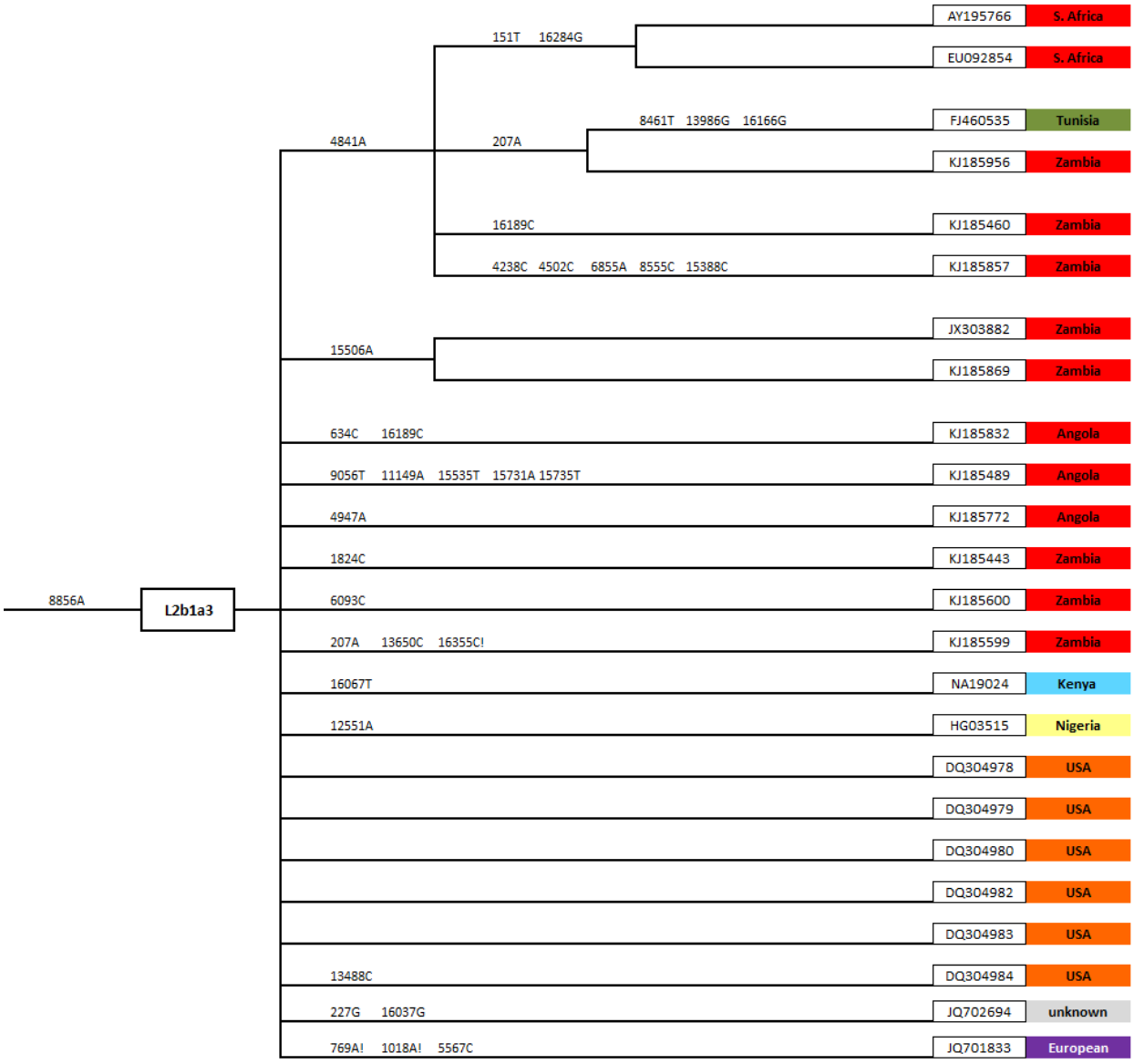

Figure 29. Phylogenetic tree of subhaplogroup L2b1a3. Polymorphic positions annotated along the branches (mutations in relation to an ancestral state), back mutations indicated with an exclamation mark. More detailed information on these samples on Annex 2.

With the increased resolution of southern clades, we discovered in some branches (L2a1a2a1a, L2a1c4a1 and L2b1a3, all <5 ka) what seems to be an association between

American samples and SA (Figure 29). In fact, together with WA, Angola and Mozambique also provided many slaves to America, mainly during the beginning of the 19th century. In 1807 the British banished slavery in their dominions and began a crusade to force the other European nations to do the same, by sending a naval squadron to the west African coast that was responsible to stop and investigate any ship suspected of carrying slaves [16].

As the number of slaves from WA decreased due to British pressure, the Portuguese tried to counterbalance this trend by sending slaves from their southern African colonies (Angola and Mozambique). High proportions of southern African, mainly Angolan, maternal ancestry in American populations has been previously reported [183]. Therefore, taking into account that L2 is present at high frequency in SA (~37 %), such a strong connection between American and southern African sequences detected in the phylogeny of L2 is not abnormal.

4.1.2.5. Other migrations

The phylogeny of L2 also harbours signs of other possible smaller migrations towards other continents. Two non-African clades stand out: L2a1k and L2a1l2a.

L2a1k is particularly curious because of its old age (~9.8 ka; ~8.8 ka). It seems to be exclusively European (Figure 30) and to have arrived in eastern Europe during the early Holocene [137, 138]. North and western African lineages are believed to have contributed to the ancient Iberian mtDNA gene pool and were possibly involved in the Post-Glacial expansions from the Franco-Cantabrian refuge area, in south-western Europe [138, 184].

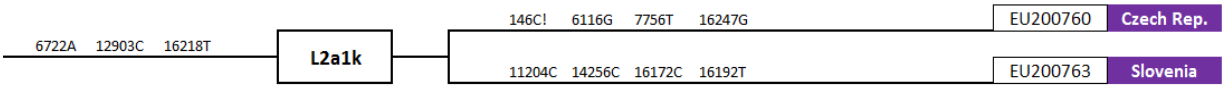

Figure 30. Phylogenetic tree of subhaplogroup L2a1k. Polymorphic positions annotated along the branches (mutations in relation to an ancestral state), back mutations indicated with an exclamation mark. More detailed information on these samples on Annex 2.

L2a1l2a harbours European and Jewish samples (Figure 31) and dates ~1.5 ka (Bayesian estimate: ~6.5 ka). Although many sequences lack additional information, taking into consideration its age and distribution, L2a1l2a is most likely related to the Ashkenazi Jewish diaspora. The Ashkenazim trace back their recent ancestry to the Rhine Valley, in central Europe, in the 1st millennium AD, but are thought to have expanded eastwards between the 11th and 15th centuries [185]. L2a1l2a has a strong eastern European component,

with samples from Romania, Poland and Russia and displays a star-like shape, the phylogenetic trademark for a recent expansion.

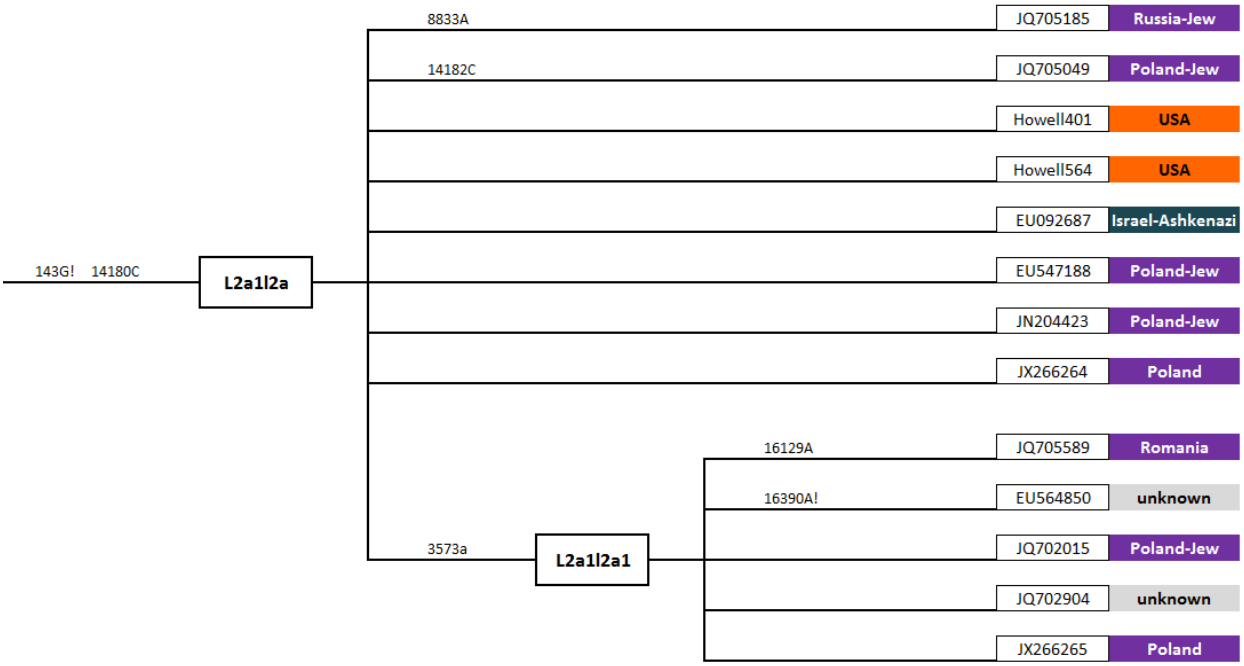

Figure 31. Phylogenetic tree of subhaplogroup L2a1l2a. Polymorphic positions annotated along the branches (mutations in relation to an ancestral state), back mutations indicated with an exclamation mark. More detailed information on these samples on Annex 2.

However, the Ashkenazi Jews are nowadays essentially present in America, due to a series of massive migrations out of eastern Europe that started in 1880, in an attempt to escape the Russian Civil War and the forced exile decreed by the Czar Nicholas I, and culminated with some of the major events of the 20th century – the Holocaust and the fall of Communism [185–187]. Regarding L2a1l2a, it also contains two samples from USA but, unfortunately, no additional information is provided. Although a recent study [140] suggests that most maternal lineages of Ashkenazi Jews trace their origin to European lineages, in this case the presence of an Ashkenazi Jew from Israel suggests an expansion from the Near East.

4.2. Expansion times

In order to assess if the movements observed in the L2 tree were accompanied by the increase in the Ne associated to the haplogroup we computed BSPs. Once again, haplogroup

L2 does not equate to population data, but a BSP applied to a specific lineage is expected to provide insights into the expansion of the population associated to that lineage – the "haplogroup-effective" population size. This kind of approach has been performed before with complete mtDNA sequences for various haplogroups with satisfying results [4, 48, 136, 140, 168].

The BSP for total sub-Saharan African L2 dataset shows two moments of increment in the Ne associated to L2 (Figure 32): ~11.5 ka, corresponding to the Pleistocene/Holocene transition and ~5 ka, probably associated to the Bantu expansion. The maximum time (~78 ka), the mean posterior estimate of the genealogy root-height, corresponds to the age of haplogroup L2.

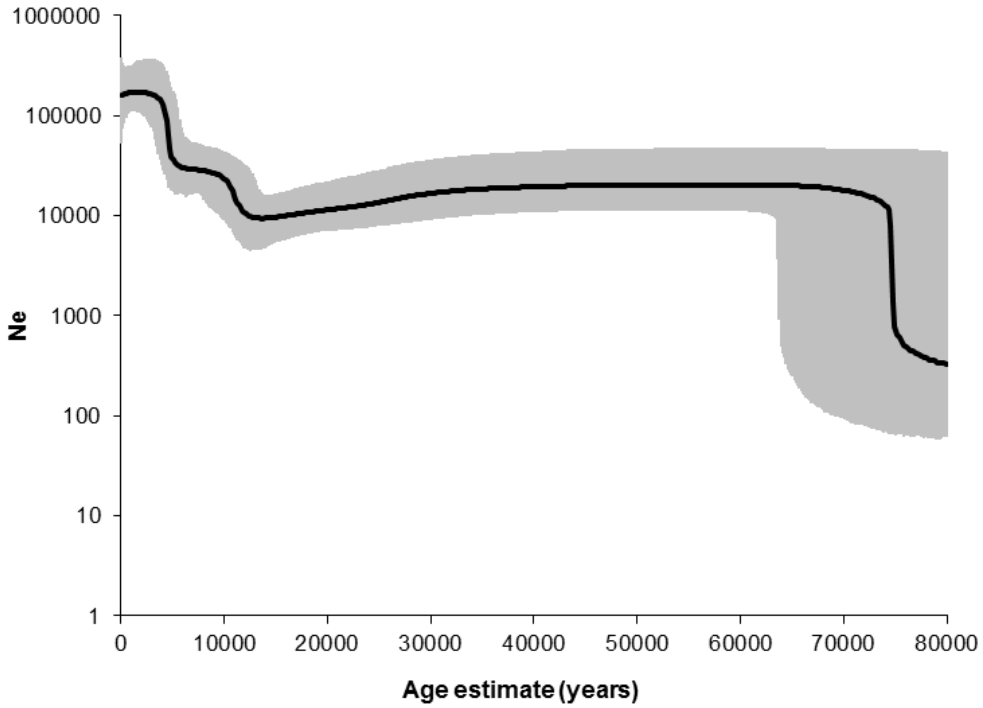

Figure 32. BSP indicating the median of the Ne associated to haplogroup L2 through time (based on all L2 complete sequences from sub-Saharan Africa). Maximum time (~78 ka) corresponds to the mean posterior estimate of the genealogy root-height.

Regional analysis allows to identify other more geographically restricted expansions (Figure 33). The BSP for Western and Central Africa (Figure 33a) clearly displays the signal of population growth during the Pleistocene/Holocene transition. The graph of increments (Figure 34) shows an additional peak ~5 ka, probably connected to the earliest steps of the Bantu migration, when Bantu-speakers were still restricted to this region.

The BSP for complete L2 east African sequences (Figure 33b) exhibits the signal of Pleistocene/Holocene transition visible in the total sub-Saharan dataset, although not very strong. Additionally, it displays another earlier and more pronounced increment (~25-30 ka), which might be connected to the signal we have detected in the phylogeny (e.g.: L2a5). L2 might have arrived to EA even earlier than we suspected, in a previous event of L2 expansion (~50 ka).

The BSP for SA (Figure 33c) shows a rapid increase in L2 Ne ~2.5 ka, consistent with the patterns observed in L2 phylogeny for the Bantu expansion (star-like shape common in the clades associated to Bantu populations dating <5 ka). This estimate for Ne increase is in agreement with what is generally accepted for the arrival of Bantu-speakers to southern Africa through the eastern coast. Bantu people are believed to have started their expansion ~5 ka (visible both in the BSP for the total African L2 dataset and for Western/Central Africa), having arrived in the Great Lakes region ~2.5 ka, expanding later into the south, having reached Mozambique by ~1.8 ka [32, 79].

The graphs of increment of the regional analysis (Figure 34) show the

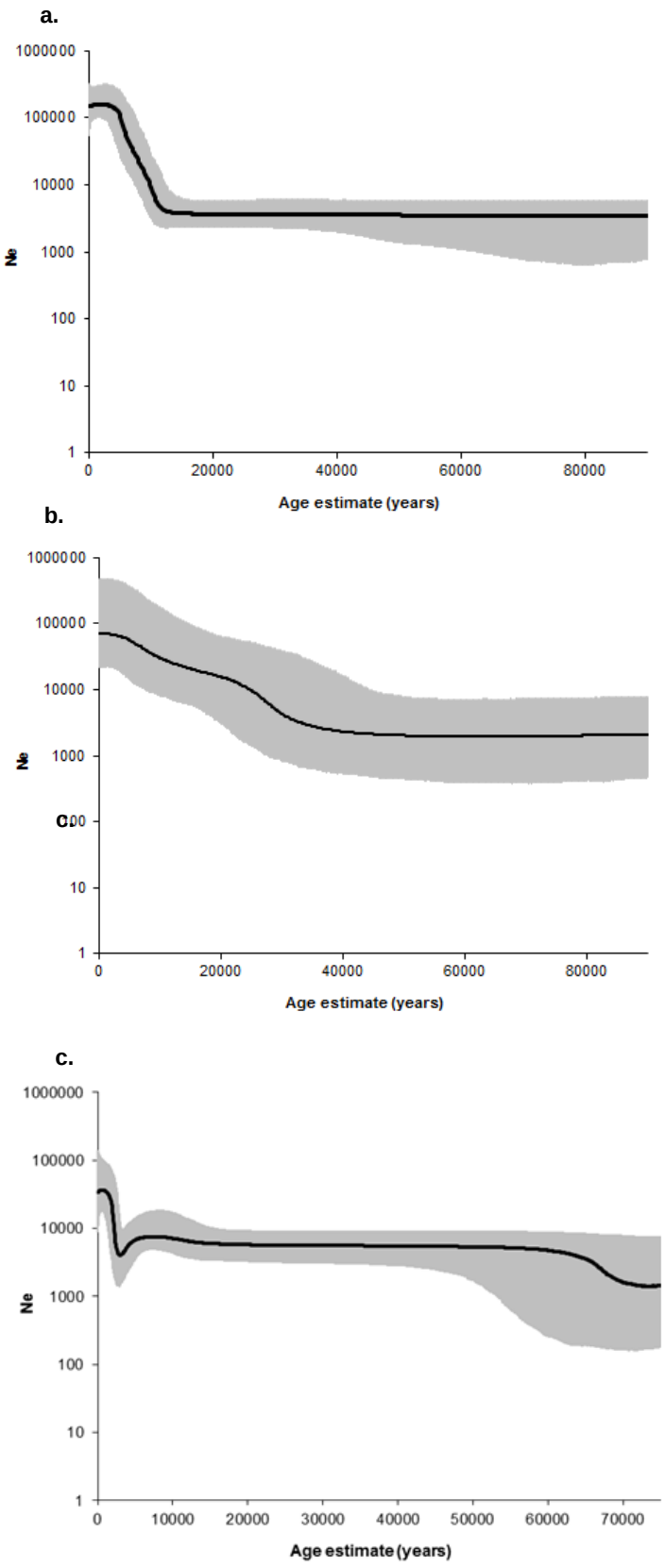

Figure 33. BSP for different African regions: Western/central Africa (a), Eastern Africa (b) and Southern Africa (c).

main variations in the Ne associated to L2 through time. All the peaks correspond to periods of known migrations in African History.

The first peak (~25-30 ka), in EA, is probably connected to the signal in L2a5 detected in the phylogenetic analysis, indicating a first expansion of L2 much earlier than initially thought, during the upper Palaeolithic. Expansions during this period were previously detected for L3b and L3d [168]. However even with the new data presented here, EA has a lower number of L2 complete samples, resulting in little precision in the graph. The second moment of L2 Ne increase in EA is almost not visible in this graph.

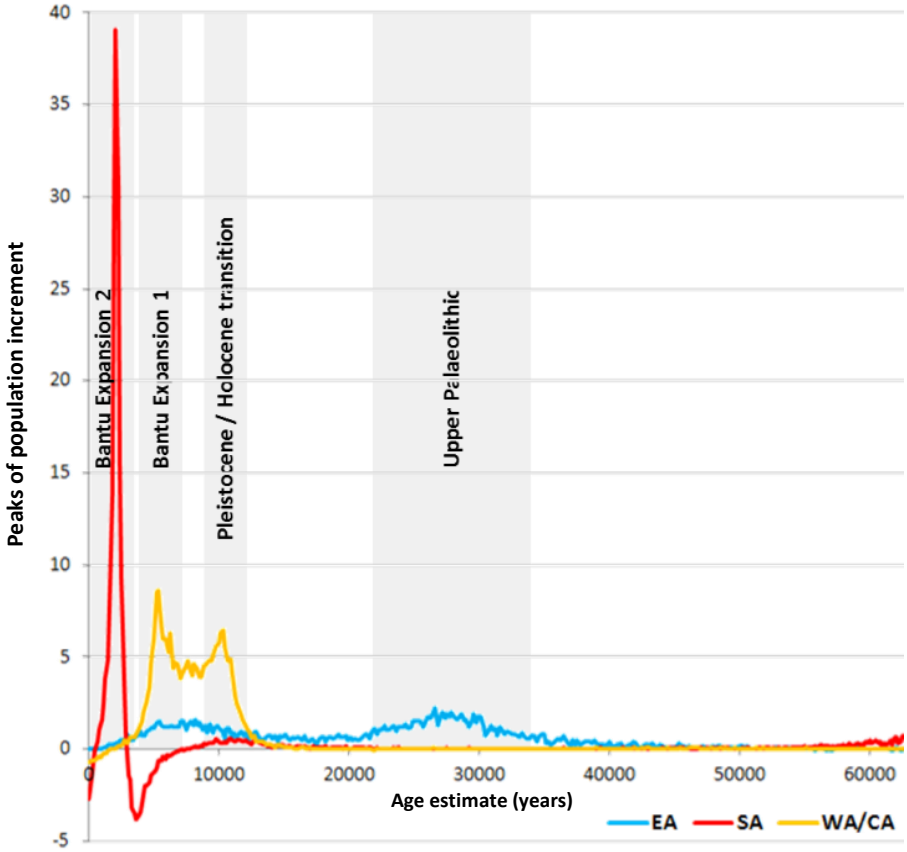

Figure 34. Graphs of population increments for EA, SA, and WA/CA.

The second increase in L2 Ne occurred ~11.5 ka, corresponding to the Pleistocene/Holocene transition, visible in the graph for Western and Central Africa. The final peaks are related to the Bantu expansion: the early stage of the migration, around 5 ka ("Bantu expansion 1"), across Central Africa, and finally, the arrival of L2 haplogroup to SA ~2 ka ("Bantu expansion 2").

4.3. Population-based analysis

As discussed before, an exclusively lineage-based analysis can be complemented with population genetics' approaches to more precisely infer the demography of populations. Therefore, in order to confirm the dynamics of Bantu and Eastern populations interactions observed within haplogroup L2, we performed a HVRI population-based analysis.

4.3.1. Genetic distances

Slatkin's linearized F_{ST} matrixes are shown in Annex S3 and S4. The MDS plots (Figure 35) have Young's S-stress values that guarantee that the plots are accurately portraying the relationship among the populations [188]. Since L2 and L0a are very common across the entire African continent (particularly in EA and SA) and might have migrated due to various independent population movements, as observed in the L2 phylogeographic analysis, excluding these lineages from the assessment of genetic distances allow to reveal some populations patterns.

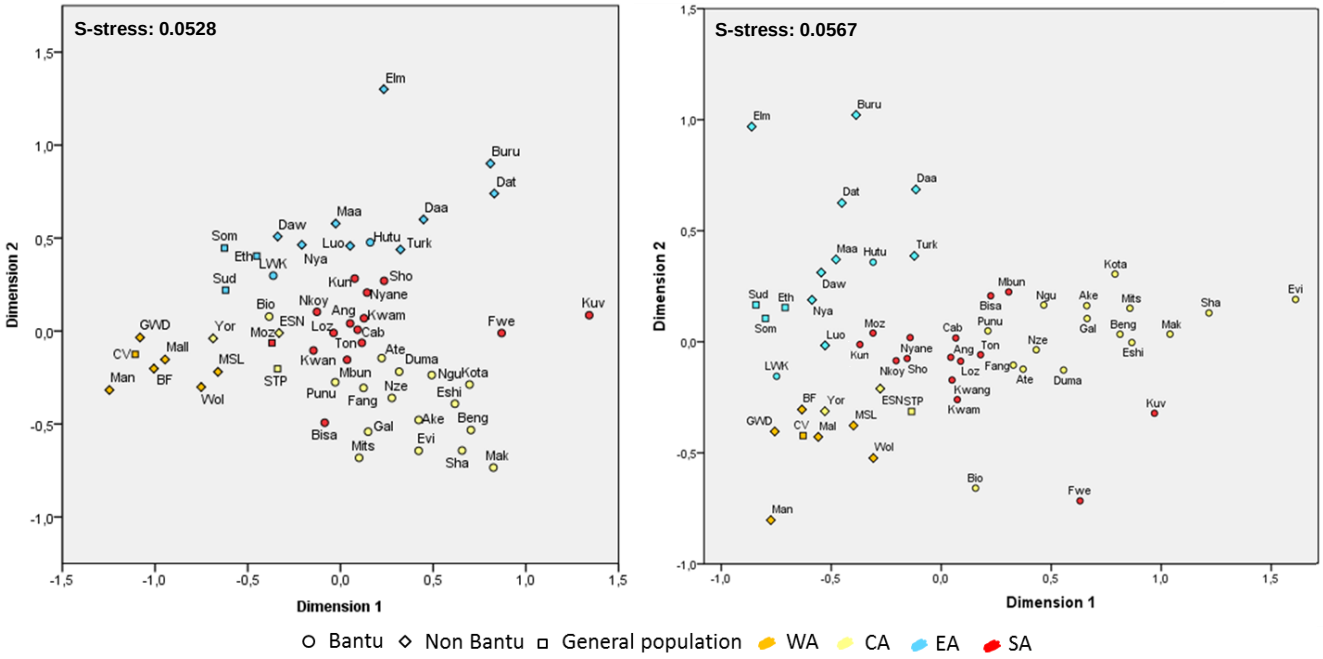

Figure 35. MDS plot based on Slatkin's linearized F_{ST}: considering all sequences (a) and excluding L2 and L0a sequences (b). Colour code as in Table 4, population code as in Annex 5.

The first dimension considering all sequences does not differentiate any comprehensive group (neither based on geography, nor on language), but separates some individual populations: the Fwe, the Kuvale (both Bantu), the Burunge and the Datog (from Tanzania). A geographic pattern is also visible, with samples from WA separated from CA, with the exceptions of Yoruba (Nigeria) and Bioko (Equatorial Guinea), which can be explained by their closest geographic proximity to WA. The majority of southern groups appear between WA and CA, which is not surprising, since southern Bantu groups derived from western and central African populations. However, there are a few exceptions (Fwe, Kuvale and Bisa).

Both the Kuvale and the Fwe, which keep their differentiated positions in the MDS excluding L2 and L0a, depart from the typical Bantu cultural standards: the Kuvale, although speaking a Bantu language, are seminomadic pastoralists [80], while the Fwe have incorporated click consonants, typical of Khoisan idioms, into their language [189]. Admixture with Khoisan neighbouring groups has been reported before for these two populations [150, 190], which might explain their distinguishable position in both plots. Bisa appears at the edge of the Bantu cluster in both plots, a result also reported elsewhere [191]. This might be due to their high frequency of L1c lineages, typical from Pygmy groups, or due to a recent arrival of a portion of the Bisa population from the Congo Basin [32, 191].

The second dimension separates in both plots EA from the other populations. The Tanzanian groups (Burunge and Datog) and El Molo, a Cushitic group from Kenya known for its genetic isolation [45], appear distant from other eastern groups in both analysis. However, when not considering L2 and L0a, eastern populations appear more scattered, and differentiation between the most peripheral groups in the first MDS (Burunge, Datog, Daasanach and El Molo) is exacerbated.

Eastern African populations appear highly dispersed in the two plots, indicating little female-mediated gene flow among them, compared to other sub-Saharan populations. This is particularly true for the Kenyan groups (El Molo, Luhya, Luo, Maasai and Turkana). Nevertheless, it is not possible to distinguish any clusters based on language or subsistence mode. On the other hand, general populations from EA (Ethiopia, Somalia and Sudan) group very close to each other. However, in the first MDS, Sudan appears in the second dimension closer to WA, due to its higher frequency of L2 [5], since this effect disappears when excluding L2 sequences from the analysis. The Daasanach, a Cushitic group from Ethiopia, appears very distant from general population and the other Ethiopian groups (Dawro-Konta and Nyangatom).

There is little evidence in the MDS of contact between Bantu and east African populations, apart from the groups near the Great Lakes. Bantu groups from this region (the Luhya from Kenya and the Hutu from Rwanda) group with EA, similarly to what was observed in previous studies [45, 192]. However, since southern Bantu populations do not cluster directly with eastern groups, this contact most likely occurred after or independently of the migration of Bantu-speakers southwards. Apart from Sudan, Luhya is the eastern population with the closest position to CA, WA and southern Bantu groups in the second dimension, probably reflecting a larger genetic ancestry of the Bantu expanding populations in this eastern Bantu group.

In Kenya, which lays on the edge of many different dispersals and harbours very differentiated ethnic groups [45], contact between Bantu and non-Bantu groups is historically recorded and supported both by linguistic and cultural evidence [193, 194]. For example, the Turkana and the Luo (both from Kenya) present typical Bantu lineages at high frequencies [45], and appear in the MDS close to a Bantu group (the Hutu). In parallel, non-Bantu eastern African lineages (L0f, L3*, L4g ,L5 and M) can be found in the Hutu mtDNA gene pool [35, 45, 192]. Curiously, the Turkana and the Luo cluster near the Hutu (from Rwanda), instead of with the Luhya, also from Kenya, which appears closer to WA when excluding L2 and L0a.

Contact between the Luo (Nilotic) and the Luhya (Bantu) resulted in the transference of some linguistic and cultural features [193, 195]. For example, although the Luo are historically pastoralists, and still preserve their strong attitudes towards cattle, they have changed to mixed farming, probably due to the influence of Bantu farmers in the region [193]. However, a complex of superiority towards the Luhya subsists among the Luo, who are extremely conservative and historically have always owned more lands and cattle than the Luhya [193]. Despite the historical and linguistic evidence, we found no indication of proximity between the Luo and the Luhya on the maternal component, which might indicate that the contact between these two groups was most likely essentially cultural.

Regarding the Luhya, their position considerably differs in both plots. Considering the entire dataset they cluster with EA but they occupy, apart from Sudan, the closest position to the CA, WA and SA Bantu groups. Excluding L2 and L0a they similarly appear in an intermediate position between EA and WA clusters. This is a signal that the Luhya, at least in their maternal component, has a dual Bantu/East African ancestry. Genetic distances give no strong evidence of contact between Bantu groups and more eastern populations, not included in the contact zone around the Great Lakes, in agreement with our phylogeography analysis. Nevertheless, South-eastern Bantus were expected to be somehow closer to eastern African

Bantus, assuming that the western Bantu route started in the region around the Great Lakes [79]. However, although Kunda (Zambia), Shona (Zimbabwe) and Nyaneka (Angola) are in some way close to the eastern cluster (including the Bantu Hutu) they are very distant from the Luhya. Moreover, in the second MDS, Kunda (Zambia) and Mozambique appear closer to the Luo from Kenya.

On the other hand, Nyaneka clusters with these south-eastern populations and not with the other Angolan populations, which are very close to Lozi, Tonga and Kwamashi (all from Zambia). All this evidence is in agreement with recent findings that suggest no differentiation in the mtDNA pool between the so-called western and eastern Bantus [74, 149] and might indicate gene flow across Bantu populations during their migration periods [80] and/or a late split between Eastern and Western groups [74].

Curiously, the general population from Mozambique (Moz) groups with populations from WA when considering the entire dataset, instead of with the other groups of South-eastern Africa. Similarly to Sudan, this might be due to a peak of frequency of L2 in this country, since in the second MDS this proximity is no longer visible.

4.3.2. Haplogroup composition

Figure 36 shows the haplogroup composition of different African regions, as well as the detailed haplogroup composition of the populations that appear in an intermediate position in the second dimension in the first MDS plot (Sudan, Luhya, Kunda, Shona and Nyaneka).

There are clear differences in mtDNA haplogroup composition between different African regions, similarly to what was reported elsewhere [25]. Western and Central Africa share the same haplogroups, although at different frequencies, with the West having the highest proportions of L2 lineages, compatible with a western origin for L2 haplogroup, and CA with higher frequency of L0a and L1 haplogroups (specifically L1c, typically associated to western Pygmy groups [152]).

EA has a differentiated haplogroup composition from the rest of sub-Saharan Africa, harbouring haplogroups that are virtually absent in other regions, such as L0*, L3*, L4 and L5, in agreement with previous results [32]. It also presents a strong non-L component, probably a result of migrations from the Arabian Peninsula back to Africa that are thought to have occurred between ~40 and 15 ka [48].

SA harbours high frequencies of L0, namely L0a, associated to the Bantu expansion [4] and L0d/L0k, specific of Khoisan populations [75]. On the other hand, eastern typical lineages are practically absent in Southern Africa, indicating that there was little mtDNA gene flow between Bantu population that migrated southwards and Eastern autochthonous groups, supporting both the phylogeographic analysis and the population results based on genetic distances.

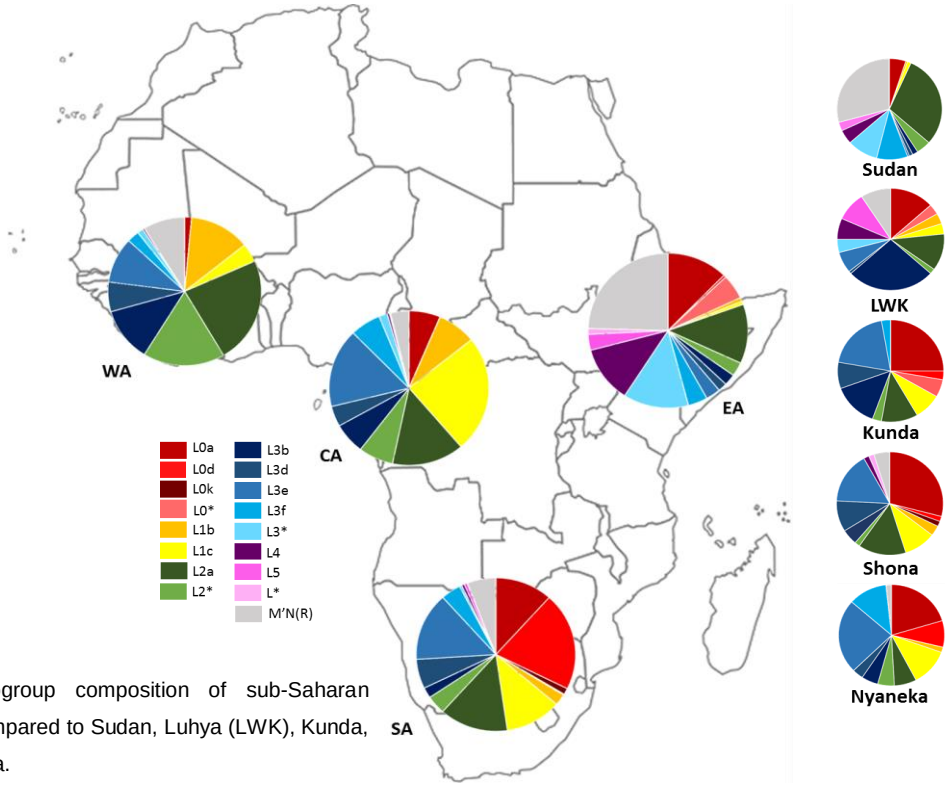

Figure 36. Haplogroup composition of sub-Saharan African regions, compared to Sudan, Luhya (LWK), Kunda, Shona and Nyaneka.

Additionally, the haplogroup composition of these five populations (Sudan, Luhya, Kunda, Shona and Nyaneka) explains their positions in the MDS plot. The frequency of L2a reaches a peak in Sudan, with a frequency (~30 %) much higher than the average for EA (~12 %) and similar to WA, whereas the proportion of L4 (~4 %), a haplogroup typically east African, is lower, which might explain its proximity to WA in the MDS. The Luhya, on the one hand, have a very high proportion of L3b (typical of WA), but, on the other hand, present L0a, L4 and L5 at frequencies comparable to EA. This might indicate that the Luhya had a high input of eastern lineages into their maternal component after arriving to the Great Lakes region.

Regarding the three southern groups, they all have a haplogroup composition similar to SA, apart from a few particularities that explain their proximity to EA in the plot. Firstly, all of them display a lower frequency of L0d compared to the general lineage composition of SA. Kunda harbours L0* (~5 %) at a frequency comparable to EA. Shona presents non-L lineages

(~5 %), together with the typically eastern L4 and L* (~3 %). Nyaneka departs from the general southern African haplogroup composition due to the frequency of L3e (~24 %) and L3f (~12 %). However, it is worth mentioning that, since these three populations have an overall haplogroup composition typically observed in other Bantu-speaking populations, their apparent proximity to EA in the first MDS is essentially due to the divergent position of Sudan and Luhya in relation to other eastern groups.

A more detailed assessment of haplogroup composition was performed, allowing to allocate specific haplogroups to certain locations (Figure 37). WA populations are relatively homogeneous, however, there seems to be a gradient in L0a frequencies from West to Central, congruent with an eastern origin of L0a and a migration carrying L0 to CA across the Sahel [4].

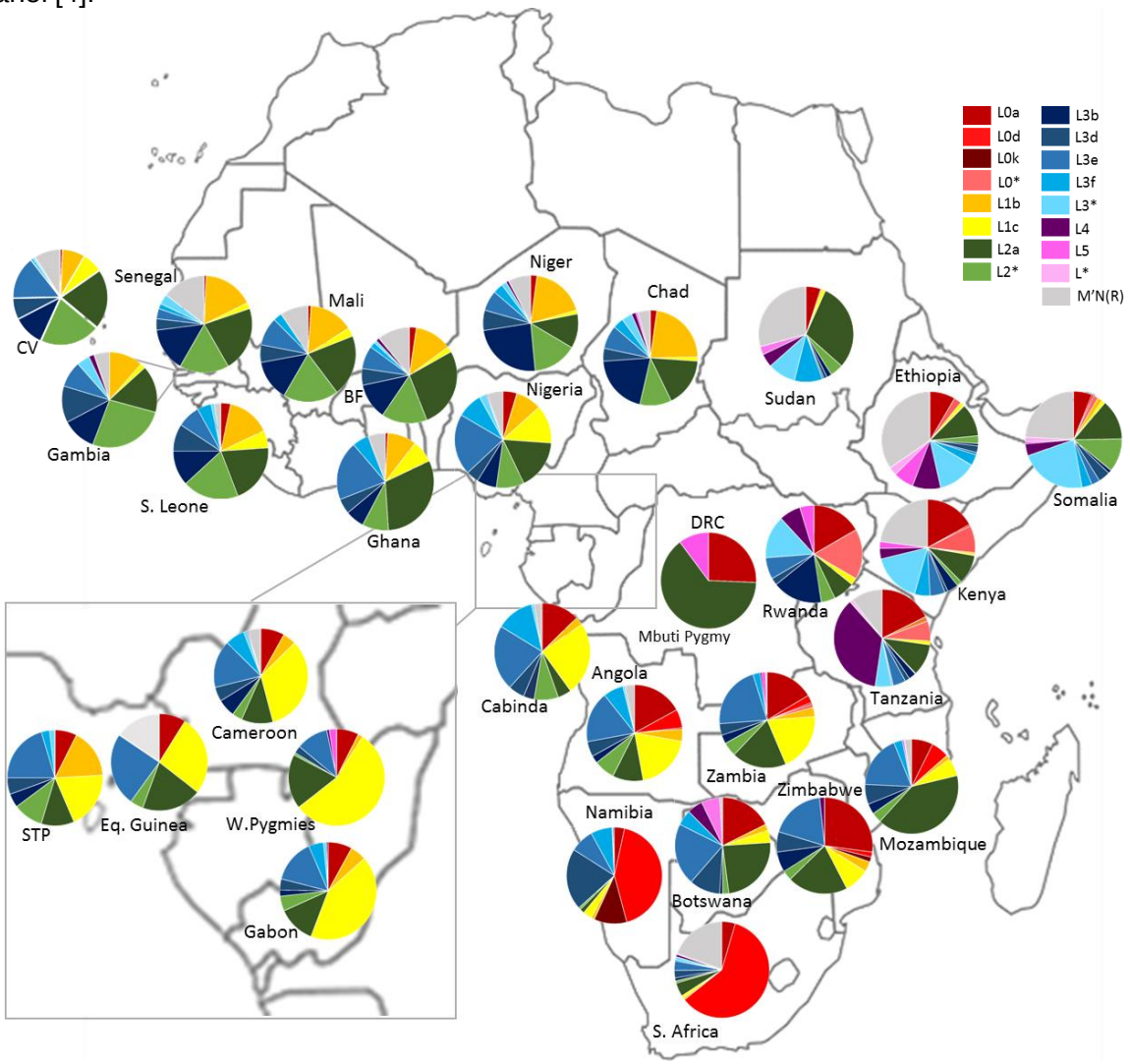

Figure 37. Haplogroup composition of sub-Saharan African countries. Population abbreviations: BF – Burkina Faso, CV – Cape Verde, S. Africa – South Africa, S. Leone – Sierra Leone, STP – São Tomé and Príncipe, W. Pygmies – Western Pygmies.

In CA there is an increase in the frequency of L0a and L2 haplogroups. Whereas in the countries closest to WA (Chad, Niger and Nigeria) L1b is very common, in the more central populations (Equatorial Guinea, Cameroon, Gabon, Western Pygmies and STP) L1c frequency is higher. Western Pygmies have a mtDNA composition very similar to the neighbouring Bantu populations, most likely due to gene flow [152]. On the contrary, the Mbuti group from DRC (Democratic Republic of Congo) stands out. This population is an eastern Pygmy group whose mtDNA gene pool has been previously described as very differentiated from western Pygmy populations [152], harbouring essentially L0a, L2a and L5 and not presenting signs of admixture with central Bantu populations.

Eastern populations closer to the Arabian Peninsula have a greater proportion of M'N(R) lineages, as expected, which are absent in Rwanda. The opposite pattern is observed for L0a, which is more common in the populations around the Great Lakes (Kenya, Rwanda and Tanzania). Tanzania has the highest proportion of L4 and a lower frequency of L3 haplogroups, whereas Sudan presents the highest frequency of L2a, as discussed above.

Regarding SA, Namibia and South Africa display high proportions of L0d and L0k, associated to the Khoisan populations that inhabit this region [75]. L0d is also present in Angola, Mozambique, Zambia and Zimbabwe at a lower frequency, but absent in Cabinda and Botswana. On the other hand, the haplogroup composition of Angola, Cabinda and Zambia is more similar to CA, indicating a central African ancestry for these populations. Botswana is the southern population with higher frequencies of L4 and L5, typical of EA, which might be a result of a migration of east African pastoralists, independent of the Bantu expansion [57], supported also by NRY analysis [196]. Mozambique has a very high proportion of L2a (as observed with the frequency distribution map – Figure 18) and L3e, which explains its odd position in the first MDS, distant from the other southern groups and closer to WA.

4.3.3. Network analysis

Pairwise genetic distances and haplogroup composition allow to infer proximities between populations, but provide no clue regarding the directionality or the age of gene flow. In this sense, a network analysis of L2 and L0a (both present in Eastern, Central and Southern Africa at considerable frequencies) provides insights on the directionality of gene flow.

Despite the loss in the resolution provided by the whole-molecule analysis, the network analysis based on HVRI has the advantage of allowing to include more individuals and populations, for which there are no complete sequences available and can be directly compared with the populations included in the MDS plots.

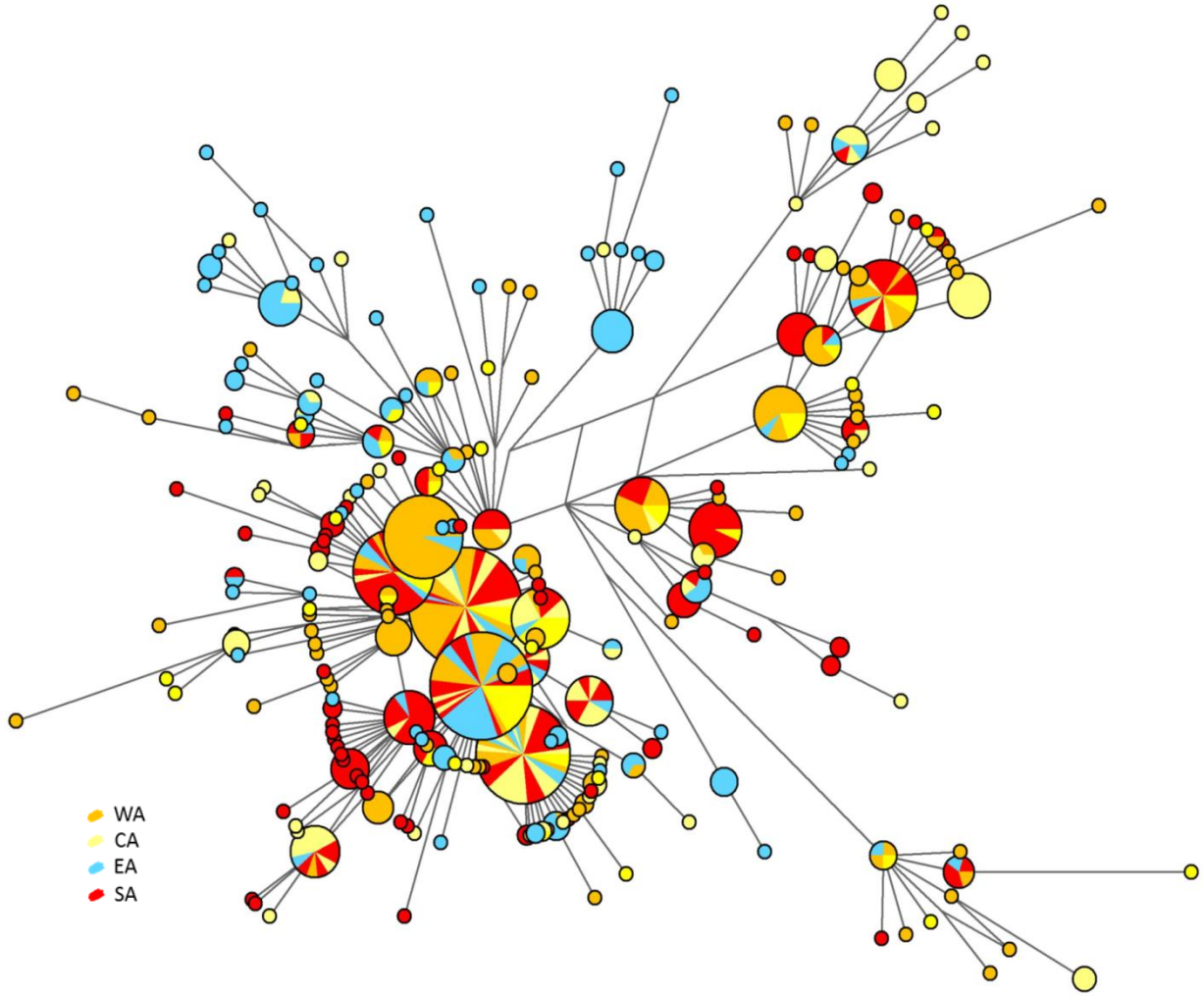

Figure 38. Reduced-median network analysis of L2 HVRI sequences included in the genetic distances analysis.

Table 6. Founder age estimates of the main L2a clades present in Eastern African populations included in the network analysis.

Clade	Founder age estimate
L2a*	15400 [1300; 29600]
L2a+16189	6500 [500; 12500]
L2a+16189+16192	10600 [0; 23600]

In both networks (Figures 38 and 39), the majority of eastern haplotypes are shared with and/or derive from WA and CA. Nevertheless, and as expected regarding our previous results, there are some shared haplotypes between eastern groups and southern Bantu populations.

Regarding L2a, the founder analysis for EA considering the three main nodes (L2a+16189, L2a+16189+16192 and L2a*) (Table 6) indicates that these haplotypes were present in EA much earlier than the Bantu expansion and, consistent with the complete sequence phylogeographic analysis, arrived in the region in the early Holocene.

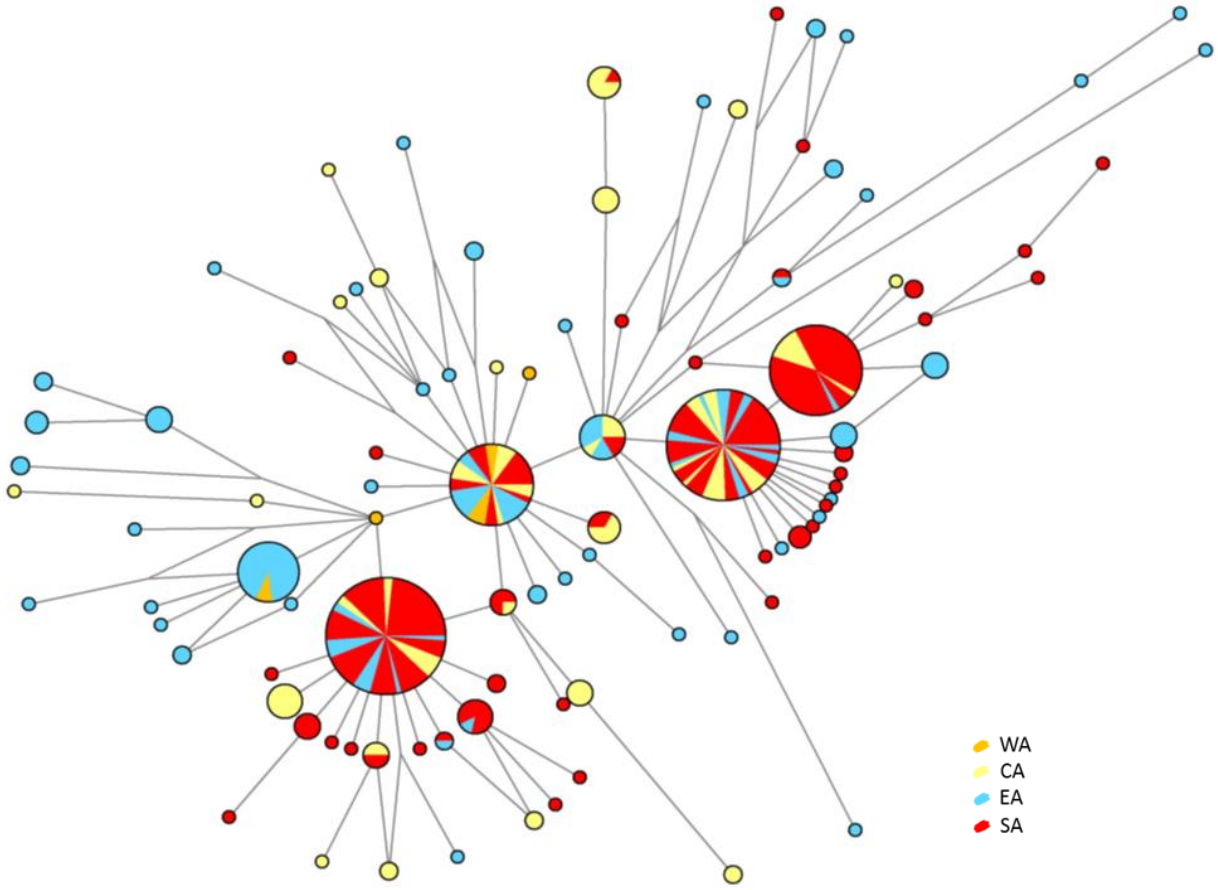

Figure 39. Reduced-median network analysis of L0a HVRI sequences included in the genetic distances analysis.

L0a founder analysis has been previously performed [4], with a peak of migration from EA to CA ~11.2 ka, consistent with the founder ages of L2a main clades from CA to EA and with our phylogenetic results (*see 4.2.1*), indicating gene flow between these regions was probably common and occurred in the early Holocene in both directions.

4.4. Additional evidence

4.4.1. Other mtDNA haplogroups

The distribution of L0a was for a long time linked to Bantu movements [161], likewise L2. However, recent evidence support an earlier expansion for L0a, probably triggered by the improvement of environmental conditions, during the Pleistocene/Holocene transition [4, 136].

The most updated version of L0a phylogeny constructed in this study (including ρ and ML age estimates for the main nodes) is shown in Annex S2. Node age estimates for the branches mentioned are shown in Table 7.

Table 7. Age estimates of L0a clades mentioned in this chapter.

Clade	ML whole mtDNA	ρ whole mtDNA	ρ synonymous age
L0a	58900 [40700 - 78000]	36600 [26400 - 47100]	3770 [20700 - 54800]
L0a1a1	13500 [8700 - 18400]	12700 [5800 - 19900]	17700 [2800 - 32700]
L0a1a2	10500 [7600 - 13300]	7800 [5500 - 10100]	6300 [4100 - 8500]
L0a1a3	13600 [8900 - 18400]	14100 [6900 - 21600]	17700 [3800 - 31700]
L0a2+16188	38100 [20800 - 56500]	29500 [15000 - 44800]	37200 [11500 - 62800]
L0a1b1a1	5200 [3300 - 7100]	4000 [2400 - 5500]	3500 [800 - 6300]
L0a2a1a	12600 [5400 - 20000]	8600 [4500 - 12800]	7900 [1800 - 14000]
L0a2a1b+16093	3500 [0 - 7600]	3500 [0 - 7600]	4300 [0 - 12700]

L0a has a pattern very similar to L2, but in an opposite direction. It had its origin in EA (after an L0 migration from the South), being nowadays also very common in CA and SA [4]. Similarly to L2, L0a moved to CA after the LGM, in the Pleistocene/Holocene transition, and was incorporated in the populations that were later involved in the Bantu migrations [4].

L0a complete sequences from EA cluster mainly with central samples (e.g.: L0a1a1, L0a1a3, L0a2+16188). There are also some branches associated with Bantu groups, such as, for instance in L0a1a2 (Figure 40), L0a2a1a, L0a2a1b+16093 and L0a1b1a1, all displaying star-like structure.

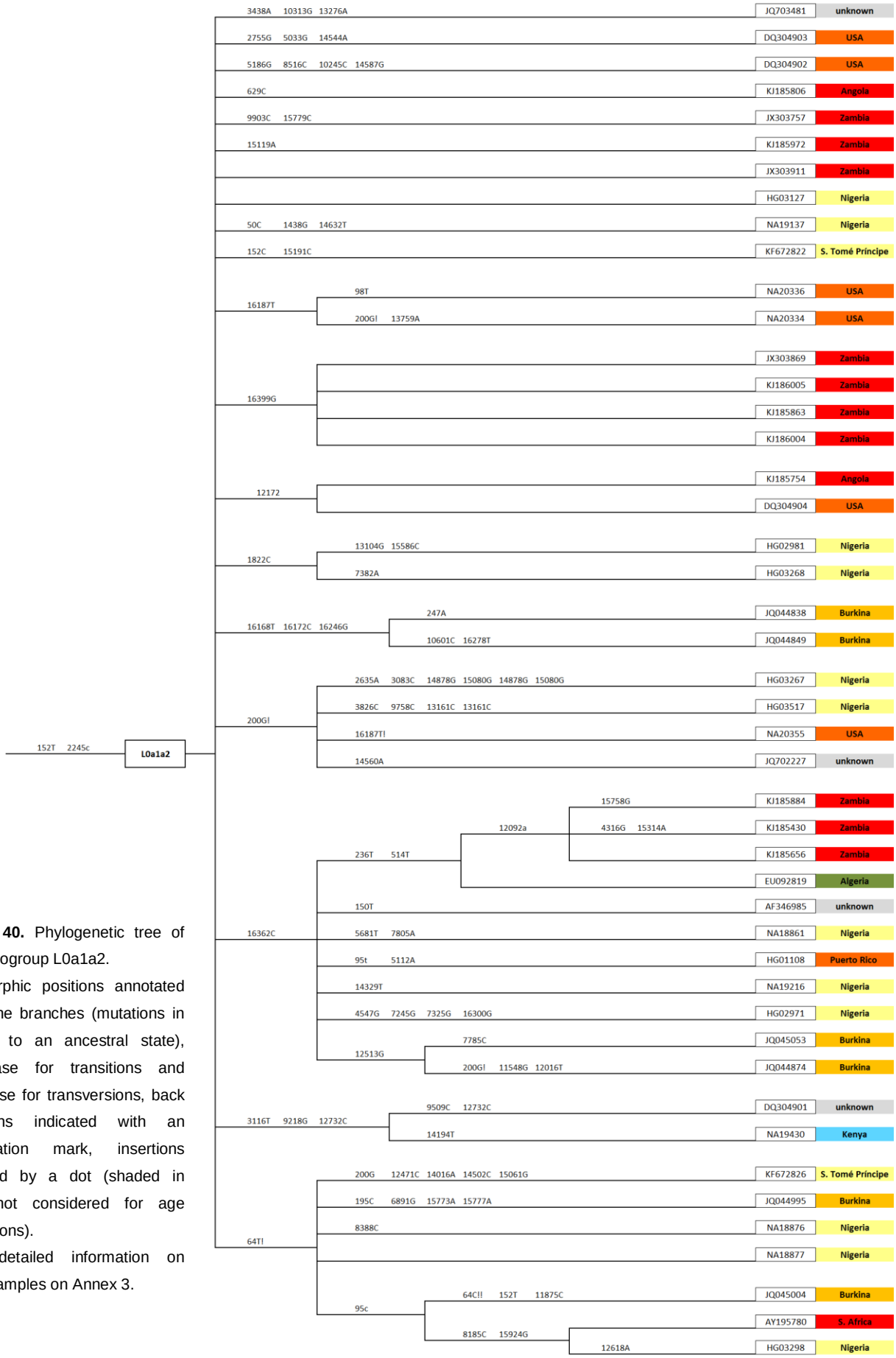

Figure 40. Phylogenetic tree of subhaplogroup L0a1a2.

Polymorphic positions annotated along the branches (mutations in relation to an ancestral state), uppercase for transitions and lowercase for transversions, back mutations indicated with an exclamation mark, insertions indicated by a dot (shaded in grey; not considered for age estimations).

More detailed information on these samples on Annex 3.

In general, similarly to what we have found for L2, southern lineages of L0a do not seem to be related with those in the East and fall within the time frame for the Bantu expansion. Additionally, Eastern typical haplogroups (L4, L5, L6, L7, L3h, L3i) are absent in SA, as previously discussed. These evidences suggest that the Bantu permanence in East Africa did not result in admixture with Eastern autochthonous populations and the population that migrated southwards had almost entirely ancestry in a west/central African mtDNA gene pool.

4.4.2. Y-chromosome

NRY allows to assess male-mediated demographic events and is, therefore, very useful to complement mtDNA results, especially in cases of sex-biased gene flow. Many variables, normally associated to the cultural background of a given population, might result in different mtDNA and NRY patterns (e.g.: patrilocality *versus* matrilocality or polygamy *versus* monogamy) and lead to asymmetrical migration and reduced male Ne [197–199].

NRY evidence indicate little gene flow in the paternal component between eastern populations and the Bantu migrants [46]. NRY Kenyan Bantu component seems to be closer to western-central populations [46], a result observed in our analysis only when excluding L2 and L0a lineages. This might be a consequence of sex-biased gene flow that resulted in more maternal lineages being shared among eastern populations (mainly L2 and L0a), regardless language affiliation. Patrilocality is highly associated to agricultural societies, such as Bantu populations [73, 199]. In this scenario, in interethnic marriages women should move to their husbands' villages, resulting in a wider dispersal of mtDNA lineages in comparison to NRY haplotypes.

A major gap in existing mtDNA studies regarding the eastern Bantu route is the lack of sampling in Uganda, which lies in the transition zone between Nilotic and Niger-Kordofanian language domains and could provide valuable insights on the dynamics of Bantu and non-Bantu populations in the Great Lakes region. On the contrary, there is NRY information available for Uganda, however only regarding Nilotes [46, 200]. The Nilotic people from Uganda present low proportion of shared NRY haplotypes with Bantu neighbours, a result consistent with a scenario of little to no gene flow between Bantu and Nilotic people.

4.4.3. Autosomal markers

Although uniparental genetic markers have been the most commonly used to assess the demographic history of populations, recent advances in sequencing technologies and data analysis software has resulted in genome-wide studies based on millions of autosomal markers scattered across the entire genome – SNPs (Single Nucleotide Polymorphisms), microsatellites (also known as STRs, Short Tandem Repeats) and/or indels. The main advantage of studying autosomal markers for demographic inference is the fact that they undergo recombination, allowing to identify and quantify admixture between populations of distinct ancestries [201].

Autosomal markers indicate admixture between eastern Nilotic and Cushitic groups [9], consistent with our results, since no differentiation between Cushites and Nilotes was noticeable in the MDS. Additionally, some groups from Sudan and Ethiopia show a considerable proportion of Near Eastern and Arabian ancestry [9]. This proximity between EA and the Arabian Peninsula was also visible in L2 and L0a trees (e.g.: clades L2a1+143+16189+16192 and L0a1d), being in agreement with previous mtDNA evidence [47, 48, 50].

There is, however some differentiated patterns between our mtDNA results and genome-wide analysis. Kenyan and Rwandan Bantus seem to have a great proportion of western and central Africa ancestry, in opposition to their maternal component, which seems to be typically Eastern [9]. The same is observed in the Luo (Nilotic group), which seems to be the only eastern non-Bantu population with such a high proportion of west/central ancestry [9]. This, together with Y-chromosome evidence [46], supports a scenario of sex-biased gene flow, consistent with agricultural societies. However, different results have been reported for populations in the Great Lakes region, especially regarding the Luhya [10, 202], probably due to differences in the number and type of markers analysed and in the populations included.

However, there is no autosomal information on south-eastern populations in order to assess whether these populations share ancestry with EA or, as observed in mtDNA variation, they mostly derive directly from central and western populations.

5. Conclusions

The present extensive phylogeographic analysis of L2 complete sequences allowed to increase the resolution of L2 phylogeny and detect branches associated to specific migrations. The increase of sampling in Eastern Africa allowed to discern population movements in which L2 was involved that were not previously known and to identify L2 lineages that were believed to be absent in this region.

Patterns observed in the phylogenetic tree, together with the Bayesian inference of variations in the L2 effective population size, indicate three main population movements associated with this haplogroup: (1) during the upper Palaeolithic, (2) post-glacial movements, probably across the Sahel and (3) the Bantu expansion.

We also found other movements in which L2 was involved, such as post-glacial movements to Europe, the transatlantic slave trade and the Ashkenazi diaspora.

Complementary population analysis confirmed the main patterns observed in the L2 phylogeny. Little evidence of mtDNA gene flow between Eastern and Southern Populations was found. The contact observed in the populations around the Great Lakes most likely occurred after or independently of the Bantu migration southwards, as evidenced both by genetic distances and haplogroup composition analysis.

6. Future perspectives

Despite our sampling, Eastern Africa remains poorly studied and more samples from this region are needed. The main gap in this area is Uganda, which lays in the transition between Niger-Kordofanian and Nilo-Saharan languages and was most likely the home of Bantu migrants before their expansion further southwards.

Nevertheless, sampling other ethnic groups, for example in Sudan and South Sudan (a region where L2 is particularly frequent), would probably provide insights into the complex migrations in Eastern Africa prior or contemporary to the arrival of Bantu-speakers.

On the other hand, whereas Southern Africa has been extensively sampled for mtDNA studies, genome-wide information regarding south-eastern populations remains unknown and would be of utmost importance to infer potential admixture of Bantu-speakers with eastern Africans. Moreover, genome-wide study on southern African Bantu groups would allow to discern patterns of ancestry and admixture between the so-called Western and Eastern Bantus, and contribute to resolve the long-lasting debate on the early Bantu migration routes.

7. References

1. Lambert CA, Tishkoff SA: **Genetic structure in African populations: implications for human demographic history**. *Cold Spring Harbor Symposia on Quantitative Biology* 2009, **74**:395–402.
2. Reed FA, Tishkoff SA: **African human diversity , origins and migrations**. *Current opinion in genetics & development* 2006, **16**(6): 597-605.
3. McDougall I, Brown FH, Fleagle JG: **Stratigraphic placement and age of modern humans from Kibish, Ethiopia.** *Nature* 2005, **433**:733–736.
4. Rito T, Richards MB, Fernandes V, Alshamali F, Cerny V, Pereira L, Soares P: **The first modern human dispersals across Africa.** *PloS one* 2013, **8**:e80031.
5. Soares P, Rito T, Pereira L, Richards M: **A genetic perspective on African prehistory**. In *Africa from MIS 6-2: Population Dynamics and Paleoenvironments*. Edited by Jones SC, Stewart B. Vertebrate Paleobiology and Paleoanthropology Book Series. Springer (in press).
6. Clark JD, Beyene Y, WoldeGabriel G, Hart WK, Renne PR, Gilbert H, Defleur A, Suwa G, Katoh S, Ludwig KR: **Stratigraphic, chronological and behavioural contexts of Pleistocene Homo sapiens from Middle Awash, Ethiopia**. *Nature* 2003, **423**:747–752.
7. Millard AR: **A critique of the chronometric evidence for hominid fossils: I. Africa and the Near East 500–50ka**. *Journal of Human Evolution* 2008, **54**:848–874.
8. Smith TM, Tafforeau P, Reid DJ, Grün R, Eggins S, Boutakiout M, Hublin J-J: **Earliest evidence of modern human life history in North African early Homo sapiens**. *Proceedings of the National Academy of Sciences* 2007, **104** (15):6128–6133.
9. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo J-M, Doumbo O, Ibrahim M, Juma AT, Kotze MJ, Lema G, Moore JH, Mortensen H, Nyambo TB, Omar SA, Powell K, Pretorius GS, Smith MW, Thera MA, Wambebe C, Weber JL, Williams SM: **The Genetic Structure and History of Africans and African Americans**. *Science* 2009, **324**:1035–1044.
10. Henn BM, Gignoux CR, Jobin M, Granka JM, Macpherson JM, Kidd JM, Rodríguez-Botigué L, Ramachandran S, Hon L, Brisbin A, Lin AA, Underhill PA, Comas D, Kidd KK,

Norman PJ, Parham P, Bustamante CD, Mountain JL, Feldman MW: **Hunter-gatherer genomic diversity suggests a southern African origin for modern humans.** *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108**:5154–62.

11. Cruciani F, Trombetta B, Massaia A, Destro-Bisol G, Sellitto D, Scozzari R: **A Revised Root for the Human Y Chromosomal Phylogenetic Tree: The Origin of Patrilineal Diversity in Africa**. *The American Journal of Human Genetics* 2011, **88**:814–818.
12. Campbell MC, Tishkoff S a: **African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping.** *Annual Review of Genomics and Human Genetics* 2008, **9**:403–33.
13. Wall JD, Lohmueller KE, Plagnol V: **Detecting ancient admixture and estimating demographic parameters in multiple human populations.** *Molecular Biology and Evolution* 2009, **26**:1823–7.
14. Mcbrearty S, Brooks AS: **The revolution that wasn't: a new interpretation of the origin of modern human behavior.** *Journal of Human Evolution* 2000, **39**:453–563.
15. Hublin JJ: **Recent human evolution in northwestern Africa.** *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 1992, **337**:185–91.
16. McEvedy C: *The Penguin Atlas of African History*. Revised ed. London: Penguin Books; 1995.
17. QEN members: Review and Atlas of Palaeovegetation: Preliminary land ecosystem maps of the world since the Last Glacial Maximum. Edited by Adams J, Faure H. Oak Ridge National Laboratory, TN, USA; 1997.
18. Nurse D, Philippson G: *The Bantu Languages*. London and New York: Routledge; 2003.
19. The International HapMap Consortium: **The International HapMap Project**. 2003, **426**:789–796.
20. Reich D, Thangaraj K, Patterson N, Price AL, Singh L: **Reconstructing Indian population history.** *Nature* 2009, **461**:489–94.
21. Manolio TA, Brooks LD, Collins FS: **A HapMap harvest of insights into the genetics of common disease.** *The Journal of Clinical Investigation* 2008, **118**:1590–605.
22. Pritchard JK, Pickrell JK, Coop G: **The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation.** *Current Biology : CB* 2010, **20**:R208–15.
23. Jallow M, Teo YY, Small KS, Rockett KA, Deloukas P, Clark TG, Kivinen K, Bojang KA, Conway DJ, Pinder M, Sirugo G, Sisay-Joof F, Usen S, Auburn S, Bumpstead SJ, Campino S, Coffey A, Dunham A, Fry AE, Green A, Gwilliam R, Hunt SE, Inouye M,

Jeffreys AE, Mendy A, Palotie A, Potter S, Ragoussis J, Rogers J, Rowlands K, et al.: **Genome-wide and fine-resolution association analysis of malaria in West Africa.** *Nature Genetics* 2009, **41**:657–65.

24. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA: **An integrated map of genetic variation from 1,092 human genomes.** *Nature* 2012, **491**:56–65.
25. Gomez F, Hirbo J, Tishkoff SA: **Genetic variation and adaptation in Africa: implications for human evolution and disease.** *Cold Spring Harbor perspectives in Biology* 2014, **6**:a008524.
26. Fage JD: **Slavery and the Slave Trade in the Context of West African History**. *The Journal of African History* 1969, **10**:393–404.
27. Salas A, Richards M, Lareu M-V, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo Á: **The African Diaspora: Mitochondrial DNA and the Atlantic Slave Trade**. *The American Journal of Human Genetics* 2004, **74**:454–465.
28. Campbell MC, Tishkoff SA: **The Evolution of Human Genetic and Phenotypic Variation in Africa**. *Current Biology* 2010, **20**:R166–R173.
29. Rosset S, Tzur S, Behar DM, Wasser WG, Skorecki K: **The population genetics of chronic kidney disease: insights from the MYH9-APOL1 locus.** *Nature Reviews Nephrology* 2011, **7**:313–26.
30. Ko W-Y, Rajan P, Gomez F, Scheinfeldt L, An P, Winkler CA, Froment A, Nyambo TB, Omar SA, Wambebe C, Ranciaro A, Hirbo JB, Tishkoff SA: **Identifying Darwinian selection acting on different human APOL1 variants among diverse African populations.** *American Journal of Human Genetics* 2013, **93**:54–66.
31. Passarino G, Semino O, Quintana-Murci L, Excoffier L, Hammer M, Santachiara-Benerecetti AS: **Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms.** *American Journal of Human Genetics* 1998, **62**:420–34.
32. Salas A, Richards M, De la Fe T, Lareu M-V, Sobrino B, Sánchez-Diz P, Macaulay V, Carracedo A: **The making of the African mtDNA landscape.** *American Journal of Human Genetics* 2002, **71**:1082–111.
33. Semino O, Santachiara-Benerecetti AS, Falaschi F, Cavalli-Sforza LL, Underhill PA: **Ethiopians and Khoisan Share the Deepest Clades of the Human Y-Chromosome Phylogeny**. *The American Journal of Human Genetics* 2002, **70**:265–268.

34. Gonder MK, Mortensen HM, Reed F a, de Sousa A, Tishkoff S a: **Whole-mtDNA genome sequence analysis of ancient African lineages.** *Molecular Biology and Evolution* 2007, **24**:757–68.

35. Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, Villems R: **Ethiopian Mitochondrial DNA Heritage: Tracking Gene Flow Across and Around the Gate of Tears**. *The American Journal of Human Genetics* 2004, **75**:752–770.

36. Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, Moral P, Dugoujon JM, Roostalu U, Loogvali EL, Kivisild T, Bandelt HJ, Richards M, Villems R, Santachiara-Benerecetti AS, Semino O, Torroni A: **The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool**. *The American Journal of Human Genetics* 2004, **75**:910–918.

37. Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, Salas A, Oppenheimer S, Macaulay V, Richards MB: **Correcting for purifying selection: an improved human mitochondrial molecular clock.** *The American Journal of Human Genetics* 2009, **84**:740–59.

38. Soares P, Alshamali F, Pereira JB, Fernandes V, Silva NM, Afonso C, Costa MD, Musilová E, Macaulay V, Richards MB, Černý V, Pereira L: **The Expansion of mtDNA Haplogroup L3 within and out of Africa**. *Molecular Biology and Evolution* 2012, **29**:915–927.

39. Greenberg J: **The classification of African languages**. *American Anthropologist* 1948, 50(1):24–30.

40. Newman J: *The Peopling of Africa: a Geographic Interpretation*. New Haven: Yale University Press; 1995.

41. Fleming HC: *Ongota: a Decisive Language in African Prehistory*. *Volume 64*. Wiesbaden: Harrassowitz; 2006.

42. Prugnolle F, Manica A, Balloux F: **Geography predicts neutral genetic diversity of human populations**. *Current Biology* 2005, **15**:R159–R160.

43. Phillipson DW: *African archaeology.* Cambridge University Press; 2005.

44. Krings M, Salem AH, Bauer K, Geisert H, Malek AK, Chaix L, Simon C, Welsby D, Di Rienzo A, Utermann G, Sajantila A, Pääbo S, Stoneking M: **mtDNA Analysis of Nile River Valley Populations: A Genetic Corridor or a Barrier to Migration?** *The American Journal of Human Genetics* 1999, **64**:1166–1176.

45. Castrí L, Garagnani P, Useli A, Pettener D, Luiselli D: **Kenyan crossroads: migration and gene flow in six ethnic groups from Eastern Africa.** *Journal of Anthropological Sciences* 2008, **86**:189–92.

46. De Filippo C, Barbieri C, Whitten M, Mpoloka SW, Gunnarsdóttir ED, Bostoen K, Nyambe T, Beyer K, Schreiber H, de Knijff P, Luiselli D, Stoneking M, Pakendorf B: **Y-chromosomal variation in sub-Saharan Africa: insights into the history of Niger-Congo groups.** *Molecular Biology and Evolution* 2011, **28**:1255–69.

47. Musilová E, Fernandes V, Silva NM, Soares P, Alshamali F, Harich N, Cherni L, Gaaied ABA El, Al-Meeri A, Pereira L, Černý V: **Population history of the Red Sea—genetic exchanges between the Arabian Peninsula and East Africa signaled in the mitochondrial DNA HV1 haplogroup**. *American Journal of Physical Anthropology* 2011, **145**:592–598.

48. Fernandes V, Alshamali F, Alves M, Costa MD, Pereira JB, Silva NM, Cherni L, Harich N, Cerny V, Soares P, Richards MB, Pereira L: **The Arabian cradle: mitochondrial relicts of the first steps along the southern route out of Africa.** *The American Journal of Human Genetics* 2012, **90**:347–55.

49. Hanotte O, Bradley DG, Ochieng JW, Verjee Y, Hill EW, Rege JEO: **African pastoralism: genetic imprints of origins and migrations.** *Science* 2002, **296**:336–9.

50. Pagani L, Kivisild T, Tarekegn A, Ekong R, Plaster C, Gallego Romero I, Ayub Q, Mehdi SQ, Thomas MG, Luiselli D, Bekele E, Bradman N, Balding DJ, Tyler-Smith C: **Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool.** *The American Journal of Human Genetics* 2012, **91**:83–96.

51. Cerný V, Pereira L, Kujanová M, Vasíková A, Hájek M, Morris M, Mulligan CJ: **Out of Arabia-the settlement of island Soqotra as revealed by mitochondrial and Y chromosome genetic diversity.** *American Journal of Physical Anthropology* 2009, **138**:439–47.

52. Segal R: *Islam's Black Slaves: A History of Africa's Other Black Diaspora*. London: Atlantic Books; 2002.

53. Badro D a, Douaihy B, Haber M, Youhanna SC, Salloum A, Ghassibe-Sabbagh M, Johnsrud B, Khazen G, Matisoo-Smith E, Soria-Hernanz DF, Wells RS, Tyler-Smith C, Platt DE, Zalloua P a: **Y-chromosome and mtDNA genetics reveal significant contrasts in affinities of modern Middle Eastern populations with European and African populations.** *PloS one* 2013, **8**:e54616.

54. Levine DN: *Greater Ethiopia: The Evolution of a Multiethnic Society*. Chicago: University of Chicago Press; 1974.

55. Richards M, Rengo C, Cruciani F, Gratrix F, Wilson JF, Scozzari R, Macaulay V, Torroni A: **Extensive female-mediated gene flow from sub-Saharan Africa into near eastern Arab populations.** *The American Journal of Human Genetics* 2003, **72**:1058–64.

56. Salas A, Richards M, De la Fe T, Lareu M-V, Sobrino B, Sánchez-Diz P, Macaulay V, Carracedo A: **The making of the African mtDNA landscape.** *The American Journal of Human Genetics* 2002, **71**:1082–111.

57. Barbieri C, Güldemann T, Naumann C, Gerlach L, Berthold F, Nakagawa H, Mpoloka SW, Stoneking M, Pakendorf B: **Unraveling the complex maternal history of Southern African Khoisan populations.** *American Journal of Physical Anthropology* 2014, **153**:435–48.

58. Güldemann T: **A linguist's view: Khoe-Kwadi speakers as the earliest food-producers of southern Africa**. *Southern African Humanities* 2008, **20**:93–132.

59. Barbujani G, Colonna V: **Human genome diversity: frequently asked questions**. *Trends in Genetics* 2010, **26**:285–295.

60. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM: **Worldwide human relationships inferred from genome-wide patterns of variation.** *Science* 2008, **319**:1100–4.

61. Schuster SC, Miller W, Ratan A, Tomsho LP, Giardine B, Kasson LR, Harris RS, Petersen DC, Zhao F, Qi J, Alkan C, Kidd JM, Sun Y, Drautz DI, Bouffard P, Muzny DM, Reid JG, Nazareth L V, Wang Q, Burhans R, Riemer C, Wittekindt NE, Moorjani P, Tindall E a, Danko CG, Teo WS, Buboltz AM, Zhang Z, Ma Q, Oosthuysen A, et al.: **Complete Khoisan and Bantu genomes from southern Africa.** *Nature* 2010, **463**:943–7.

62. Hassan FA: *Droughts, Food and Culture. Ecological Change and Food Security in Africa's Later Prehistory*. New York: Plenum Publishers; 2002.

63. Nicoll K: **Recent environmental change and prehistoric human activity in Egypt and Northern Sudan**. *Quaternary Science Reviews* 2004, **23**:561–580.

64. Jousse H: **What is the impact of Holocene climatic changes on human societies? Analysis of West African Neolithic populations dietary customs**. *Quaternary International* 2006, **151**:63–73.

65. Kuper R: **Climate-Controlled Holocene Occupation in the Sahara: Motor of Africa's Evolution**. *Science* 2006, **313**:803–807.

66. Brooks N, Chiapello I, Lernia S Di, Drake N, Legrand M, Moulin C, Prospero J: **The climate-environment-society nexus in the Sahara from prehistoric times to the present day**. *The Journal of North African Studies* 2005, **10**:253–292.

67. Yokoyama Y, Lambeck K, De Deckker P, Johnston P, Fifield L: **Timing of the Last Glacial Maximum from observed sea-level minima**. *Nature* 2000, **406**:713–6.

68. Ghienne J-F, Schuster M, Bernard A, Duringer P, Brunet M: **The Holocene giant Lake Chad revealed by digital elevation models**. *Quaternary International* 2002, **87**:81–85.

69. Prentice IC, Jolly D: **Mid-Holocene and glacial-maximum vegetation geography of the northern continents and Africa**. *Journal of Biogeography* 2000, **27**:507–519.

70. Jolly D, Harrison S., Damnati B, Bonnefille R: **Simulated climate and biomes of Africa during the late quaternary**. *Quaternary Science Reviews* 1998, **17**:629–657.

71. Al-Abri A, Podgorná E, Rose JI, Pereira L, Mulligan CJ, Silva NM, Bayoumi R, Soares P, Cerný V: **Pleistocene-Holocene boundary in Southern Arabia from the perspective of human mtDNA variation.** *American Journal of Physical Anthropology* 2012, **149**:291–8.

72. Soares P, Achilli A, Semino O, Davies W, Macaulay V, Bandelt H-J, Torroni A, Richards MB: **The archaeogenetics of Europe.** *Current Biology* 2010, **20**:R174–83.

73. Pakendorf B, de Filippo C, Bostoen K: **Molecular Perspectives on the Bantu Expansion: A Synthesis**. *Language Dynamics and Change* 2011, **1**:50–88.

74. De Filippo C, Bostoen K, Stoneking M, Pakendorf B: **Bringing together linguistic and genetic evidence to test the Bantu expansion.** *Proceedings of the Royal Society B: Biological Sciences* 2012, **279**:3256–63.

75. Barbieri C, Vicente M, Rocha J, Mpoloka SW, Stoneking M, Pakendorf B: **Ancient Substructure in Early mtDNA Lineages of Southern Africa**. *The American Journal of Human Genetics* 2013, **92**(2): 285-292..

76. Thomas MG, Parfitt T, Weiss D a, Skorecki K, Wilson JF, le Roux M, Bradman N, Goldstein DB: **Y chromosomes traveling south: the cohen modal haplotype and the origins of the Lemba – the "Black Jews of Southern Africa".** *The American Journal of Human Genetics* 2000, **66**:674–86.

77. Zhivotovsky L a, Underhill P a, Cinnioğlu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, Chambers GK, Herrera RJ, Yong KK, Gresham D, Tournev I, Feldman MW, Kalaydjieva L: **The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time.** *The American Journal of Human Genetics* 2004, **74**:50–61.

78. Berniell-Lee G, Calafell F, Bosch E, Heyer E, Sica L, Mouguiama-Daouda P, van der Veen L, Hombert J-M, Quintana-Murci L, Comas D: **Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages.** *Molecular Biology and Evolution* 2009, **26**:1581–9.

79. Diamond J, Bellwood P: **Farmers and their languages: the first expansions.** *Science* 2003, **300**:597–603.

80. Alves I, Coelho M, Gignoux C, Damasceno A, Prista A, Rocha J: **Genetic Homogeneity Across Bantu-Speaking Groups from Mozambique and Angola Challenges Early Split Scenarios between East and West Bantu Populations**. *Human Biology* 2011, **83**:13–38.

81. Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE, Shriver MD: **Estimating African American admixture proportions by use of population-specific alleles.** *The American Journal of Human Genetics* 1998, **63**:1839–51.

82. Tyner J, Houston D: **Controlling Bodies: The Punishment of Multiracialized Sexual Relations**. *Antipode* 2000, **32**:387–409.

83. Lind JM, Hutcheson-Dilks HB, Williams SM, Moore JH, Essex M, Ruiz-Pesini E, Wallace DC, Tishkoff S a, O'Brien SJ, Smith MW: **Elevated male European and female African contributions to the genomes of African American individuals.** *Human Genetics* 2007, **120**:713–22.

84. Alves-Silva J, da Silva Santos M, Guimarães PE, Ferreira a C, Bandelt HJ, Pena SD, Prado VF: **The ancestry of Brazilian mtDNA lineages.** *The American Journal of Human Genetics* 2000, **67**:444–61.

85. Amorim A: **Archaeogenetics**. *Journal of Iberian Archaeology* 1999:15–26.

86. Menozzi P, Piazza A, Cavalli-Sforza L: **Synthetic maps of human gene frequencies in Europeans**. *Science* 1978, **201**:786–792.

87. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H: **Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction**. *Cold Spring Harbor Symposia on Quantitative Biology* 1986, **51**:263–273.

88. Sanger F, Nicklen S, Coulson AR: **DNA sequencing with chain-terminating inhibitors**. *Proceedings of the National Academy of Sciences* 1977, **74**:5463–5467.

89. Sanger F, Coulson AR: **A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase**. *Journal of Molecular Biology* 1975, **94**:441–448.

90. Underhill PA, Kivisild T: **Use of y chromosome and mitochondrial DNA population structure in tracing human migrations**. *Annual Review of Genetics* 2007, **41**:539–564.

91. Avise J, Arnold J, Ball R: **Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics**. *Annual Review of Ecology and Systematics* 1987:489–522.

92. Hewitt GM: **Speciation, hybrid zones and phylogeography—or seeing genes in space and time**. *Molecular Ecology* 2001, **10**:537–549.

93. Cavalli-Sforza LL, Feldman MW: **The application of molecular genetic approaches to the study of human evolution.** *Nature Genetics* 2003, **33 Suppl**(march):266–75.

94. Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG: **Sequence and organization of the human mitochondrial genome**. *Nature* 1981, **290**:457–465.

95. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N: **Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA**. *Nature Genetics* 1999, **23**:147.

96. Fernández-Silva P, Enriquez J, Montoya J: **Replication and transcription of mammalian mitochondrial DNA**. *Experimental Physiology* 2003, **88**:41–56.

97. Falkenberg M, Larsson N-G, Gustafsson CM: **DNA replication and transcription in mammalian mitochondria.** *Annual Review of Biochemistry* 2007, **76**:679–99.

98. Brown TA, Cecconi C, Tkachuk AN, Bustamante C, Clayton DA: **Replication of mitochondrial DNA occurs by strand displacement with alternative light-strand origins, not via a strand-coupled mechanism.** *Genes & Development* 2005, **19**:2466–76.

99. Shadel GS, Clayton DA: **Mitochondrial dna maintenance in vertebrates**. *Annual Review of Biochemistry* 1997, **66**(1), 409-435..

100. Brown TA, Clayton DA: **Release of replication termination controls mitochondrial DNA copy number after depletion with 2',3'-dideoxycytidine**. *Nucleic Acids Research* 2002, **30**:2004–2010.

101. Barrell BG, Bankier AT, Drouin J: **A different genetic code in human mitochondria**. *Nature* 1979, **282**:189–194.

102. Pakendorf B, Stoneking M: **Mitochondrial DNA and human evolution**. *Annual Review of Genomics and Human Genetics* 2005, **6**:165–183.

103. Brown WM, George M, Wilson AC: **Rapid evolution of animal mitochondrial DNA.** *Proceedings of the National Academy of Sciences of the United States of America* 1979, **76**:1967–1971.

104. Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC: **Natural selection shaped regional mtDNA variation in humans.** *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**:171–6.

105. Soares P, Abrantes D, Rito T, Thomson N, Radivojac P, Li B, Macaulay V, Samuels DC, Pereira L: **Evaluating purifying selection in the mitochondrial DNA of various mammalian species**. *PloS one* 2013, **8**:e58993.

106. Shuster RC, Rubenstein AJ, Wallace DC: **Mitochondrial DNA in anucleate human blood cells**. *Biochemical and Biophysical Research Communications* 1988, **155**:1360–1365.

107. Wiesner RJ, Rüegg JC, Morano I: **Counting target molecules by exponential polymerase chain reaction: Copy number of mitochondrial DNA in rat tissues**. *Biochemical and Biophysical Research Communications* 1992, **183**:553–559.

108. Lopez J V., Yuhki N, Masuda R, Modi W, O'Brien SJ: **Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat**. *Journal of Molecular Evolution* 1994, **39**:174–190.

109. Hazkani-Covo E, Zeller RM, Martin W: **Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes.** *PLoS Genetics* 2010, **6**:e1000834.

110. **MITOMAP: A Human Mitochondrial Genome Database** (http://www.mitomap.org).

111. Cann RL, Stoneking M, Wilson AC: **Mitochondrial DNA and human evolution.** *Nature* 1987, **325**:31–6.

112. Tang H, Siegmund DO, Shen P, Oefner PJ, Feldman MW: **Frequentist Estimation of Coalescence Times From Nucleotide Sequence Data Using a Tree-Based Partition**. *Genetics* 2002, **161**:447–459.

113. Ingman M, Kaessmann H, Pääbo S, Gyllensten U: **Mitochondrial genome variation and the origin of modern humans.** *Nature* 2000, **408**:708–13.

114. Richards M, Côrte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, Hedges R, Bandelt HJ, Sykes B: **Paleolithic and neolithic lineages in the European mitochondrial gene pool**. *The American Journal of Human Genetics* 1996, **59**(1): 185.

115. Destro-Bisol G, Jobling MA, Rocha J, Novembre J, Richards MB, Mulligan C, Batini C, Manni F: **Molecular anthropology in the genomic era**. *Journal of Anthropological Sciences* 2010, **88**:93–112.

116. Torroni A, Schurr TG, Cabell MF, Brown MD, Neel J V, Larsen M, Smith DG, Vullo CM, Wallace DC: **Asian affinities and continental radiation of the four founding Native American mtDNAs**. *The American Journal of Human Genetics* 1993, **53**:563–590.

117. Richards M: **Tracing European Founder Lineages in the Near Eastern mtDNA Pool**. *The American Journal of Human Genetics* 2000, **67**:1251–1276.

118. Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonne-Tamir B, Sykes B, Torroni A: **The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs**. *The American Journal of Human Genetics* 1999, **64**:232–249.

119. Torroni A, Huoponen K, Francalacci P: **Classification of European mtDNAs from an analysis of three European populations**. *Genetics* 1996, **144**:1835–1850.

120. Templeton AR: **Human origins and analysis of mitochondrial DNA sequences**. *Science* 1992, **255**:737.

121. Excoffier L, Langaney A: **Origin and differentiation of human mitochondrial DNA**. *The American Journal of Human Genetics* 1989, **44**(1): 73..

122. Van Oven M, Kayser M: **Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation**. *Human Mutation* 2009, **30**:E386–94.

123. Torroni A, Achilli A, Macaulay V, Richards M, Bandelt H-J: **Harvesting the fruit of the human mtDNA tree.** *Trends in Genetics* 2006, **22**:339–45.

124. Harpending H, Rogers A: **Genetic perspectives on human origins and differentiation.** *Annual Review of Genomics and Human Genetics* 2000, **1**:361–85.

125. Harpending HC, Batzer MA, Gurven M, Jorde LB, Rogers AR, Sherry ST: **Genetic traces of ancient demography**. *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**:1961–1967.

126. Behar DM, van Oven M, Rosset S, Metspalu M, Loogväli E-L, Silva NM, Kivisild T, Torroni A, Villems R: **A "Copernican" reassessment of the human mitochondrial DNA tree from its root.** *The American Journal of Human Genetics* 2012, **90**:675–84.

127. Elson JL, Turnbull DM, Howell N: **Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection.** *The American Journal of Human Genetics* 2004, **74**:229–38.

128. Harich N, Costa M, Fernandes V, Kandil M, Pereira J, Silva N, Pereira L: **The trans-Saharan slave trade - clues from interpolation analyses and high-resolution characterization of mitochondrial DNA lineages**. *BMC Evolutionary Biology* 2010, **10**:138.

129. Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, Calderon FL, Simionati B, Valle G, Richards M, Macaulay V, Scozzari R: **Do the Four Clades of the mtDNA Haplogroup L2 Evolve at Different Rates?** *The American Journal of Human Genetics* 2001, **69**:1348–1356.

130. Watson E, Forster P, Richards M, Bandelt H: **Mitochondrial footprints of human expansions in Africa**. *The American Journal of Human Genetics* 1997:691–704.

131. Coia V, Destro-Bisol G, Verginelli F, Battaggia C, Boschi I, Cruciani F, Spedini G, Comas D, Calafell F: **Brief communication: mtDNA variation in North Cameroon: lack of Asian lineages and implications for back migration from Asia to sub-Saharan Africa.** *The American Journal of Human Genetics* 2005, **128**:678–81.

132. Quintana-Murci L, Chaix R, Wells R: **Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor**. *The American Journal of Human Genetics* 2004:827–845.

133. Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ, Amorim A: **Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade**. *Annals of Human Genetics* 2001, **65**:439–458.

134. Rosa A, Brehem A: **African human mtDNA phylogeography at-a-glance.** *Journal of Anthropological Sciences* 2011, **89**:25–58.

135. Basell L: **Middle Stone Age (MSA) site distributions in eastern Africa and their relationship to Quaternary environmental change, refugia and the evolution of Homo sapiens**. *Quaternary Science Reviews* 2008, **27**:2484–2498.

136. Atkinson QD, Gray RD, Drummond AJ: **Bayesian coalescent inference of major human mitochondrial DNA haplogroup expansions in Africa.** *Proceedings of the Royal Society B: Biological Sciences* 2009, **276**:367–73.

137. Cerezo M, Achilli A, Olivieri A, Perego U a, Gómez-Carballa A, Brisighelli F, Lancioni H, Woodward SR, López-Soto M, Carracedo A, Capelli C, Torroni A, Salas A: **Reconstructing ancient mitochondrial DNA links between Africa and Europe.** *Genome Research* 2012, **22**:821–6.

138. Malyarchuk BA, Derenko M, Perkova M, Grzybowski T, Vanecek T, Lazur J: **Reconstructing the phylogeny of African mitochondrial DNA lineages in Slavs.** *European Journal of Human Genetics* 2008, **16**:1091–6.

139. Mielnik-Sikorska M, Daca P, Malyarchuk B: **The History of Slavs Inferred from Complete Mitochondrial Genome Sequences**. *PloS one* 2013, **8**:e54360.

140. Costa MD, Pereira JB, Pala M, Fernandes V, Olivieri A, Achilli A, Perego UA, Rychkov S, Naumova O, Hatina J, Woodward SR, Eng KK, Macaulay V, Carr M, Soares P, Pereira L, Richards MB: **A substantial prehistoric European ancestry amongst Ashkenazi maternal lineages.** *Nature Communications* 2013, **4**:2543.

141. Mairal Q, Santos C, Silva M, Marques SL, Ramos A, Aluja MP, Amorim A, Prata MJ, Alvarez L: **Linguistic isolates in Portugal: insights from the mitochondrial DNA pattern.** *Forensic Science International Genetics* 2013, **7**:618–23.

142. Alvarez L, Santos C, Ramos A, Pratdesaba R, Francalacci P, Aluja MP: **Mitochondrial DNA patterns in the Iberian Northern plateau: population dynamics and substructure of the Zamora province**. *American Journal of Physical Anthropology* 2010, **142**:531–539.

143. Pereira L, Prata MJ, Amorim A: **Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation**. *Annals of Human Genetics* 2000, **64**:491–506.

144. Gonzalez AM, Brehm A, Perez JA, Maca-Meyer N, Flores C, Cabrera VM: **Mitochondrial DNA affinities at the Atlantic fringe of Europe**. *American Journal of Physical Anthropology* 2003, **120**:391–404.

145. Santos C, Fregel R, Cabrera VM, Alvarez L, Larruga JM, Ramos A, López MA, Pilar Aluja M, González AM: **Mitochondrial DNA and Y-chromosome structure at the Mediterranean and Atlantic façades of the Iberian Peninsula.** *American Journal of Human Biology* 2014, **26**:130–41.

146. Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbujani G: **Reconstruction of prehistory on the basis of genetic data.** *The American Journal of Human Genetics* 2000, **66**:1177–9.

147. Bandelt H-J, Lahermo P, Richards M, Macaulay V: **Detecting errors in mtDNA data by phylogenetic analysis**. *International Journal of Legal Medicine* 2001, **115**:64–69.

148. Langaney A, Roessli D, Blyenburgh NH, Dard P: **Do most human populations descend from phylogenetic trees?** *Human Evolution* 1992, **7**:47–61.

149. Barbieri C, Vicente M, Oliveira S, Bostoen K, Rocha J, Stoneking M, Pakendorf B: **Migration and Interaction in a Contact Zone: mtDNA Variation among Bantu-Speakers in Southern Africa.** *PloS one* 2014, **9**:e99117.

150. Barbieri C, Butthof A, Bostoen K, Pakendorf B: **Genetic perspectives on the origin of clicks in Bantu languages from southwestern Zambia.** *European Journal of Human Genetics* 2013, **21**:430–6.

151. Barbieri C, Whitten M, Beyer K, Schreiber H, Li M, Pakendorf B: **Contrasting maternal and paternal histories in the linguistic context of Burkina Faso.** *Molecular Biology and Evolution* 2012, **29**:1213–23.

152. Batini C, Lopes J, Behar DM, Calafell F, Jorde LB, van der Veen L, Quintana-Murci L, Spedini G, Destro-Bisol G, Comas D: **Insights into the demographic history of African Pygmies from complete mitochondrial genomes.** *Molecular Biology and Evolution* 2011, **28**:1099–110.

153. Maca-Meyer N, Gonzalez A, Larruga J, Flores C, Cabrera V: **Major genomic mitochondrial lineages delineate early human expansions**. *BMC Genetics* 2001, **2**:13.

154. Sambrook J, Fritsch E, Maniatis T: *Molecular Cloning: a Laboratory Manual*. 2nd edition. New York: Cold Spring Harbor Laboratory Press; 1989.

155. Walsh P, Metzger D, Higuchi R: **Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material.** *Biotechniques* 1991, **54**:134–139.

156. Drummond A, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A, Ashton B: **Geneious v5.4**. 2011.

157. Kloss-Brandstätter A, Pacher D, Schönherr S, Weissensteiner H, Binna R, Specht G, Kronenberg F: **HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups**. *Human Mutatation* 2011, **32**:25–32.

158. Cerezo M, Achilli A, Olivieri A: **Reconstructing ancient mitochondrial DNA links between Africa and Europe**. *Genome Research* 2012:821–826.

159. Derenko M, Malyarchuk B, Bahmanimehr A, Denisova G, Perkova M, Farjadian S, Yepiskoposyan L: **Complete Mitochondrial DNA Diversity in Iranians**. *PLoS ONE* 2013, **8**:e80673.

160. Pereira L, Freitas F, Fernandes V, Pereira JB, Costa MD, Costa S, Máximo V, Macaulay V, Rocha R, Samuels DC: **The diversity present in 5140 human mitochondrial genomes.** *The American Journal of Human Genetics* 2009, **84**:628–40.

161. Bandelt H-J, Forster P, Sykes BC, Richards MB: **Mitochondrial portraits of human populations using median networks.** *Genetics* 1995, **141**:743–753.

162. Forster P, Harding R, Torroni a, Bandelt HJ: **Origin and evolution of Native American mtDNA variation: a reappraisal.** *The American Journal of Human Genetics* 1996, **59**:935–45.

163. Saillard J, Forster P, Lynnerup N, Nørby S: **mtDNA Variation among Greenland Eskimos: The Edge of the Beringian Expansion**. *The American Journal of Human Genetics* 2000, **1**:718–726.

164. Yang Z: **PAML: a program package for phylogenetic analysis by maximum likelihood**. *Computer Applications in the Biosciences* 1997, **13** (5):555–556.

165. Hasegawa M, Kishino H, Yano T: **Dating of the human-ape splitting by a molecular clock of mitochondrial DNA**. *Journal of Molecular Evolution* 1985, **22**:160–174.

166. Howell N, Elson JL, Turnbull DM, Herrnstadt C: **African Haplogroup L mtDNA sequences show violations of clock-lfike evolution.** *Molecular Biology and Evolution* 2004, **21**:1843–54.

167. Drummond AJ, Suchard MA, Xie D, Rambaut A: **Bayesian phylogenetics with BEAUti and the BEAST 1.7.** *Molecular Biology and Evolution* 2012, **29**:1969–73.

168. Soares P, Alshamali F, Pereira JB, Fernandes V, Silva NM, Afonso C, Costa MD, Musilová E, Macaulay V, Richards MB, Cerny V, Pereira L: **The Expansion of mtDNA Haplogroup L3 within and out of Africa.** *Molecular Biology and Evolution* 2012, **29**:915–27.

169. Kass RE, Raftery AE: **Bayes Factors**. *Journal of the American Statistical Association* 2012, **90**:773–795.

170. Drummond AJ, Rambaut A, Shapiro B, Pybus OG: **Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences**. *Molecular Biology and Evolution* 2005, **22**:1185–1192.

171. Rambaut A, Suchard M, Xie D, Drummond A: **Tracer v1.6, Available from http://beast.bio.ed.ac.uk/Tracer**. 2014.

172. Drummond A, Rambaut A: **BEAST: Bayesian evolutionary analysis by sampling trees**. *BMC Evolutionary Biololgy* 2007, **7**:214.

173. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A: **Relaxed phylogenetics and dating with confidence.** *PLoS Biology* 2006, **4**:e88.

174. Fagundes NJR, Kanitz R, Eckert R, Valls ACS, Bogo MR, Salzano FM, Smith DG, Jr. WAS, Zago MA, Ribeiro-dos-Santos AK, Santos SEB, Petzl-Erler ML, Bonatto SL: **Mitochondrial Population Genomics Supports a Single Pre-Clovis Origin with a**

Coastal Route for the Peopling of the Americas. *The American Journal of Human Genetics* 2008, **82**:583–592.

175. Batini C, Lopes J, Behar DM, Calafell F, Jorde LB, van der Veen L, Quintana-Murci L, Spedini G, Destro-Bisol G, Comas D: **Insights into the demographic history of African Pygmies from complete mitochondrial genomes.** *Molecular Biology and Evolution* 2011, **28**:1099–110.
176. Excoffier L, Lischer HEL: **Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows.** *Molecular Ecology Resources* 2010, **10**:564–7.
177. Slatkin M: **A Measure of Population Subdivision Based on Microsatellite Allele Frequencies**. *Genetics* 1995, **139**:457–462.
178. Endicott P, Ho SYW, Metspalu M, Stringer C: **Evaluating the mitochondrial timescale of human evolution.** *Trends in Ecology & Evolution* 2009, **24**:515–21.
179. Rando JC, Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM, Bandelt H-J: **Mitochondrial DNA analysis of Northwest African populations reveals genetic exchanges with European, Near-Eastern, and sub-Saharan populations**. *Annals of Human Genetics* 1998, **62**:531–550.
180. Brehm A, Pereira L, Bandelt H, Prata MJ, Amorim A: **Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade**. 2002 **66** (1):49–60.
181. Scholz CA, Johnson TC, Cohen AS, King JW, Peck JA, Overpeck JT, Talbot MR, Brown ET, Kalindekafe L, Amoako PYO, Lyons RP, Shanahan TM, Castañeda IS, Heil CW, Forman SL, McHargue LR, Beuning KR, Gomez J, Pierson J: **East African megadroughts between 135 and 75 thousand years ago and bearing on early-modern human origins.** *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**:16416–21.
182. Scheinfeldt LB, Soi S, Tishkoff SA: **Working toward a synthesis of archaeological, linguistic, and genetic data for inferring African population history**. *Proceedings of the National Academy of Sciences* 2010, **107**(Supplement 2):8931–8938.
183. Stefflova K, Dulik MC, Barnholtz-Sloan JS, Pai A a, Walker AH, Rebbeck TR: **Dissecting the within-Africa ancestry of populations of African descent in the Americas.** *PloS one* 2011, **6**:e14495.
184. Pereira L, Cerny V, Cerezo M, Silva NM, Hajek M, Vasikova A, Kujanova M, Brdicka R, Salas A: **Linking the sub-Saharan and West Eurasian gene pools: maternal and**

paternal heritage of the Tuareg nomads from the African Sahel. *European Journal of Human Genetics* 2010, **18**:915–923.

185. Markel H: **Di Goldine Medina (The Golden Land): Historical Perspectives of Eugenics and the East European (Ashkenazi) Jewish-American Community, 1880-1925**. *Health Matrix* 1997, **7**:49.

186. Budnitskiĭ O V.: **Jews, Pogroms, and the White Movement: A Historiographical Critique**. *Kritika: Explorations in Russian and Eurasian History* 2001, **2**:751–772.

187. Klier J: *Russians, Jews, and the Pogroms of 1881-1882*. Cambridge University Press; 2011.

188. Sturrock K, Rocha J: **A Multidimensional Scaling Stress Evaluation Table**. *Field Methods* 2000, **12**:49–60.

189. Bostoen K, Sands B: **Clicks in south-western Bantu languages: contact-induced vs. language-internal lexical change**. *Proceedings of the 6th World Congress of African Linguistics* 2012, **3**:45.

190. Coelho M, Sequeira F, Luiselli D, Beleza S, Rocha J: **On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola**. *BMC Evolutionary Biology* 2009, **9**:80.

191. De Filippo C, Heyn P, Barham L, Stoneking M, Pakendorf B: **Genetic perspectives on forager-farmer interaction in the Luangwa valley of Zambia.** *American Journal of Physical Anthropology* 2010, **141**:382–94.

192. Castrì L, Tofanelli S, Garagnani P, Bini C, Fosella X, Pelotti S, Paoli G, Pettener D, Luiselli D: **mtDNA variability in two Bantu-speaking populations (Shona and Hutu) from Eastern Africa: implications for peopling and migration patterns in sub-Saharan Africa.** *American Journal of Physical Anthropology* 2009, **140**:302–11.

193. Odak O: **Inter-ethnic relations in bantu-nilotic ethnic boundaries of western kenya**. *Zeitschrift für Ethnologie* 1995:227–240.

194. Oliver R: **The nilotic contribution to bantu africa**. *The Journal of African History* 1982, **23**:433–442.

195. Dimmendaal GJ: **Language Ecology and Linguistic Diversity on the African Continent**. *Language and Linguistics Compass* 2008, **2**:840–858.

196. Henn BM, Gignoux C, Lin A a, Oefner PJ, Shen P, Scozzari R, Cruciani F, Tishkoff S a, Mountain JL, Underhill P a: **Y-chromosomal evidence of a pastoralist migration through Tanzania to southern Africa.** *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**:10693–8.

197. Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M: **Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence.** *Nature Genetics* 2001, **29**:20–1.

198. Pilkington MM, Wilder J a, Mendez FL, Cox MP, Woerner A, Angui T, Kingan S, Mobasher Z, Batini C, Destro-Bisol G, Soodyall H, Strassmann BI, Hammer MF: **Contrasting signatures of population growth for mitochondrial DNA and Y chromosomes among human populations in Africa.** *Molecular Biology and Evolution* 2008, **25**:517–25.

199. Wilkins JF, Marlowe FW: **Sex-biased migration in humans: what should we expect from genetic data?** *Bioessays* 2006, **28**:290–300.

200. Gusmão L, Sánchez-Diz P, Gomes I, Alves C, Carracedo Á, João Prata M, Amorim A: **Genetic analysis of autosomal and Y-specific STRs in the Karimojong population from Uganda**. *International Congress Series* 2006, **1288**:213–215.

201. Novembre J, Ramachandran S: **Perspectives on human population structure at the cusp of the sequencing era**. *Annual Review of Genomics and Human Genetics* 2011, **12**:245–274.

202. Sikora M, Laayouni H, Calafell F, Comas D, Bertranpetit J: **A genomic analysis identifies a novel component in the genetic structure of sub-Saharan African populations.** *European Journal of Human Genetics* 2011, **19**:84–8.

8. Annexes

Annexes in digital support

Annex S1. Phylogenetic tree of mtDNA haplogroup L2.

Annex S2. Phylogenetic tree of mtDNA haplogroup L0a.

Annex S3. Slatkin's linearized FST matrix.

Annex S3. Slatkin's linearized FST matrix (excluding L0a and L2).

Annex 1. List of primers used in amplification and sequencing reactions [153].

L and H stand for Light- and Heavy-strand, respectively.

Fragment	Primer (position)	Primer sequence (5'-3')	Fragment size (bp)	Fragment	Primer (position)	Primer sequence (5'-3')	Fragment size (bp)
1	P1F-L16340	AGCCATTTACCGTACATAGCACA	681	**11**	P11F-L5278	TGGGCCATTATCGAAGAATT	593
	P1R-H408	TGTTAAAAGTGCATACCGCCA			P11R-H5832	GACAGGGGTTAGGCCTCTTT	
2	P2F-L382	CAAAGAACCCTAACACCAGCC	603	**12**	P12F-L5781	AGCCCCGGCAGGTTTGAAGC	626
	P2R-H945	GGGAGGGGGTGATCTAAAAC			P12R-H6367	TGGCCCCTAAGATAGAGGAGA	
3	P3F-L923	GTCACACGATTAACCCAAGTCA	607	**13**	P13F-L6337	CCTGGAGCCTCCGTAGACCT	601
	P3R-H1487	GTATACTTGAGGAGGGTGACGG			P13R-H6899	GCACTGCAGCAGATCATTTC	
4	P4F-L1466	GAGTGCTTAGTTGAACAGGGCC	629	**14**	P14F- L6869	CCGGCGTCAAAGTATTTAGC	578
	P4R-H2053	TTAGAGGGTTCTGTGGGCAAA			P14R-H7406	GGGTTCTTCGAATGTGTGGTAG	
5	P5F-L2025	GCCTGGTGATAGCTGGTTGTCC	609	**15**	P15F-L7379	AGAAGAACCCTCCATAAACCTG	580
	P5R-H2591	GGAACAAGTGATTATGCTACCT			P15R-H7918	AGATTAGTCCGCCGTAGTCG	
6	P6F-L2559	CACCGCCTGCCCAGTGACACAT	591	**16**	P16F-L7882	TCCCTCCCTTACCATCAAATCA	506
	P6R-H3108	TCGTACAGGGAGGAATTTGAA			P16R-H8345	TTTCACTGTAAAGAGGTGTTGG	
7	P7F-L3073	AAAGTCCTACGTGATCTGAGTTC	640	**17**	P17F-L8299	ACCCCCTCTAGAGCCCACTG	603
	P7R-H3670	GGCGTAGTTTGAGTTTGATGC			P17R-H8861	GAGCGAAAGCCTATAATCACTG	
8	P8F-L3644	GCCACCTCTAGCCTAGCCGT	623	**18**	P18F-L8799	CTCGGACTCCTGCCTCACTCA	638
	P8R-H4227	ATGCTGGAGATTGTAATGGGT			P18R-H9397	GTGGCCTTGGTATGTGCTTT	
9	P9F-L4210	CCACTCACCCTAGCATTACTTA	625	**19**	P19F-L9362	GGCCTACTAACCAACACACTA	609
	P9R-H4792	ACTCAGAAGTGAAAGGGGGCTA			P19R-H9928	AACCACATCTACAAAATGCCAGT	
10	P10F-L4750	CCAATACTACCAATCAATACTC	599	**20**	P20F-L9886	TCCGCCAACTAATATTTCACTT	617
	P10R-H5306	GGTGATGGTGGCTATGATGGTG			P20R-H10462	AATGAGGGGCATTTGGTAAA	

Annex 1. *(continued).*

Fragment	Primer (position)	Primer sequence (5'-3')	Fragment size (bp)	Fragment	Primer (position)	Primer sequence (5'-3')	Fragment size (bp)
21	P21F-L10403	AAAGGATTAGACTGAACCGAA	612	**27**	P27F-L13612	AAGCGCCTATAGCACTCGAA	614
	P21R-H10975	CCATGATTGTGAGGGGTAGG			P27R-H14186	TGGTTGAACATTGTTTGTTGG	
22	P22F-L10949	CTCCGACCCCCTAACAACCC	617	**28**	P28F-L14125	TCTTTCTTCTTCCCACTCATCC	602
	P22R-H11527	CAAGGAAGGGGTAGGCTATG			P28R-H14685	CATTGGTCGTGGTTGTAGTCC	
23	P23F-L11486	AAAACTAGGCGGCTATGGTA	629	**29**	P29F-L14650	CCCCATTACTAAACCCACACTC	604
	P23R-H12076	GGAGAATGGGGGATAGGTGT			P29R-H15211	TTGAACTAGGTCTGTCCCAATG	
24	P24F-L12028	GGCTCACTCACCCACCACATT	615	**30**	P30F-L15162	CTCCCGTGAGGCCAAATATC	597
	P24R-H12603	ACGAACAATGCTACAGGGATG			P30R-H15720	GTCTGCGGCTAGGAGTCAAT	
25	P25F-L12572	ACAACCCAGCTCTCCCTAAG	591	**31**	P31F-L15676	TCCCCATCCTCCATATATCC	524
	P25R-H13124	ATTTTCTGCTAGGGGGTGGA			P31R-H16157	TGATGTGGATTGGGTTTTTATGTA	
26	P26F-L13088	AGCCCTACTCCACTCAAGCAC	618	**32**	P32F-L15996	CTCCACCATTAGCACCCAAAGC	446
	P26R-H13666	AGGGTGGGGTTATTTTCGTT			P32R-H16401	TGATTTCACGGAGGATGGTG	

Annex 2. List of 801 complete L2 sequences used for phylogenetic reconstruction and phylogeography analysis.

Codes for regions as in Table 1. Abbreviations: G. Bissau – Guinea-Bissau, S. Africa – South Africa, STP – São Tomé and Príncipe, S. Arabia – Saudi Arabia, USA – United States of America. Additional information on ethnicity or geographical location indicated whenever provided by the authors.

Sample	Haplogroup	Region	Country / Ethnic group	Additional information	Reference
ETH31	L2a1+143	EA	Ethiopia	-	Present study
EU092671	L2a1+143+16189	NA	Morocco	Jew	[1]
HG03121	L2a1+143+16189	CA	Nigeria	Esan	[2]
JQ044837	L2a1+143+16189	WA	Burkina Faso	-	[3]
SUD43	L2a1+143+16189	EA	Sudan	-	Present study
SUD12	L2a1+143+16189	EA	Sudan	-	Present study
SOM49	L2a1+143+16189	EA	Somalia	-	Present study
ETH5	L2a1+143+16189	EA	Ethiopia	-	Present study
SOM76	L2a1+143+16189	EA	Somalia	-	Present study
EU092679	L2a1+143+16189+16192	AP / NE	Israel	Palestinian	[1]
EU092782	L2a1+143+16189+16192	AP / NE	Oman	-	[1]
EU092793	L2a1+143+16189+16192	AP / NE	Yemen	-	[1]
JQ702430	L2a1+143+16189+16192	AM	USA	-	[1]
EU092806	L2a1+143+16189+16192	NA	Morocco	-	[1]
Howell434	L2a1+143+16189+16192	AM	USA	-	[4]
EU092659	L2a1+143+16189+16192	AP / NE	Israel	Druze	[1]
FJ460527	L2a1+143+16189+16192	NA	Tunisia	-	[5]
EU092674	L2a1+143+16189+16192	EA	Ethiopia	Jew	[1]
HG03342	L2a1+143+16189+16192	CA	Nigeria	Esan	[2]
ETH23	L2a1+143+16189+16192	EA	Ethiopia	-	Present study
EU092823	L2a1+143+16189+16192	NA	Libya	-	[1]
EU092658	L2a1+143+16189+16192	AP / NE	Israel	Bedouin	[1]
EU597491	L2a1+143+16189+16192	AP / NE	Israel	Bedouin	[6]
FJ460520	L2a1+143+16189+16309	NA	Tunisia	-	[5]
JQ044822	L2a1+143+16189+16309	WA	Burkina Faso	-	[3]
JQ044839	L2a1+143+16189+16309	WA	Burkina Faso	-	[3]
HG03485	L2a1+143+16189+16309	WA	Sierra Leone	Mende	[2]
SOM17	L2a1+143+16189+16309	EA	Somalia	-	Present study
SOM59	L2a1+143+16189+16309	EA	Somalia	-	Present study
SOM96	L2a1+143+16189+16309	EA	Somalia	-	Present study
SOM81	L2a1+143+16189+16309	EA	Somalia	-	Present study
JQ045062	L2a1+16189	WA	Burkina Faso	-	[3]
HG03028	L2a1+16189	WA	Gambia	-	[2]
Howell195	L2a1a	AM	USA	-	[4]
Howell382	L2a1a	AM	USA	-	[4]
Howell577	L2a1a	AM	USA	-	[4]
EU092916	L2a1a	EA	Kenya	-	[4]
DQ304926	L2a1a	AM	USA	-	[7]

Annex 2. *(continued)*

JQ044922	L2a1a	WA	Burkina Faso	-	[3]
DQ304927	L2a1a	AM	USA	-	[7]
HG03114	L2a1a	CA	Nigeria	Esan	[2]
DQ304928	L2a1a	AM	USA	-	[7]
HG02588	L2a1a	WA	Gambia	-	[2]
DQ304924	L2a1a	AM	USA	-	[7]
DQ304932	L2a1a	AM	USA	-	[7]
HG02501	L2a1a	AM	Barbados	African Caribbean	[2]
JQ044942	L2a1a	WA	Burkina Faso	-	[3]
JQ044885	L2a1a	WA	Burkina Faso	-	[3]
JQ044961	L2a1a	WA	Burkina Faso	-	[3]
SUD56	L2a1a	EA	Sudan	-	Present study
HG02852	L2a1a	WA	Gambia	-	[2]
HG02860	L2a1a	WA	Gambia	-	[2]
JQ044927	L2a1a	WA	Burkina Faso	-	[3]
JQ044983	L2a1a	WA	Burkina Faso	-	[3]
JQ044992	L2a1a	WA	Burkina Faso	-	[3]
JQ705087	L2a1a	-	unknown	-	[8]
JQ044911	L2a1a	WA	Burkina Faso	-	[3]
JQ044818	L2a1a	WA	Burkina Faso	-	[3]
HG03563	L2a1a	WA	Sierra Leone	Mende	[2]
KJ185692	L2a1a	SA	Zambia	Mbunda	[9]
KJ185688	L2a1a	SA	Zambia	Mbunda	[9]
MOZ326	L2a1a	SA	Mozambique	-	Present study
MOZ33	L2a1a	SA	Mozambique	-	Present study
SOM68	L2a1a	EA	Somalia	-	Present study
Howell162	L2a1a1	AM	USA	-	[4]
Howell156	L2a1a1	AM	USA	-	[4]
Howell571	L2a1a1	AM	USA	-	[4]
DQ304925	L2a1a1	AM	USA	-	[7]
DQ304933	L2a1a1	AM	USA	-	[7]
JQ706014	L2a1a1	-	unknown	-	[8]
JQ045001	L2a1a1	WA	Burkina Faso	-	[3]
JQ044977	L2a1a1	WA	Burkina Faso	-	[3]
JQ045005	L2a1a1	WA	Burkina Faso	-	[3]
JQ045020	L2a1a1	WA	Burkina Faso	-	[3]
JQ045066	L2a1a1	WA	Burkina Faso	-	[3]
JQ045074	L2a1a1	WA	Burkina Faso	-	[3]
KJ185589	L2a1a1	SA	Zambia	Lozi	[9]
KJ185982	L2a1a1	SA	Zambia	Lunda	[9]
KJ185827	L2a1a1	SA	Angola	Ovimbundu	[9]
HG03460	L2a1a1	WA	Sierra Leone	Mende	[2]
HG03578	L2a1a1	WA	Sierra Leone	Mende	[2]
HG02922	L2a1a1	CA	Nigeria	Esan	[2]
HG03410	L2a1a1	WA	Sierra Leone	Mende	[2]
HG03548	L2a1a1	WA	Sierra Leone	Mende	[2]

Annex 2. *(continued)*

HQ425645	L2a1a2	AM	USA	-	Family Tree
HG01363	L2a1a2	AM	Colombia	Medellín	[2]
JQ044987	L2a1a2	WA	Burkina Faso	-	[3]
HG03123	L2a1a2	CA	Nigeria	Esan	[2]
KC622072	L2a1a2	SA	Botswana	Kalanga	[10]
JQ044997	L2a1a2	WA	Burkina Faso	-	[3]
JQ044809	L2a1a2	WA	Burkina Faso	-	[3]
JQ044812	L2a1a2	WA	Burkina Faso	-	[3]
JQ044952	L2a1a2	WA	Burkina Faso	-	[3]
JQ045076	L2a1a2	WA	Burkina Faso	-	[3]
JQ045027	L2a1a2	WA	Burkina Faso	-	[3]
JQ044879	L2a1a2	WA	Burkina Faso	-	[3]
JQ044884	L2a1a2	WA	Burkina Faso	-	[3]
JQ044945	L2a1a2	WA	Burkina Faso	-	[3]
JQ044951	L2a1a2	WA	Burkina Faso	-	[3]
JQ045048	L2a1a2	WA	Burkina Faso	-	[3]
Howell570	L2a1a2	AM	USA	-	[4]
DQ304971	L2a1a2	AM	USA	-	[7]
KC622070	L2a1a2	SA	Botswana	Tswana	[10]
KJ185440	L2a1a2	SA	Zambia	Luvale	[9]
MOZ88	L2a1a2	SA	Mozambique	-	Present Study
JN214449	L2a1a2	EUR	Italy	Umbria (Centre)	[11]
JN214457	L2a1a2	EUR	Italy	Marche (Centre)	[11]
Howell569	L2a1a2	AM	USA	-	[4]
Howell380	L2a1a2	AM	USA	-	[4]
DQ304977	L2a1a2	AM	USA	-	[7]
DQ304968	L2a1a2	AM	USA	-	[7]
AF346977	L2a1a2	-	unknown	Effik	[12]
EU092691	L2a1a2	SA	Mozambique	Tswa	[1]
EU092804	L2a1a2	AP / NE	Yemen	-	[1]
EU092933	L2a1a2	AS	Pakistan	Makrani	[1]
JQ705250	L2a1a2	-	unknown	-	[8]
KC622069	L2a1a2	SA	Botswana	Tswana	[10]
KJ185769	L2a1a2	SA	Angola	Nyaneka	[9]
JX303760	L2a1a2	SA	Zambia	-	[13]
KJ185896	L2a1a2	SA	Zambia	Kwangwa	[9]
KJ185768	L2a1a2	SA	Angola	Nyaneka	[9]
KJ185829	L2a1a2	SA	Angola	Ovimbundu	[9]
KJ185934	L2a1a2	SA	Zambia	Makoma	[9]
MOZ142	L2a1a2	SA	Mozambique	-	Present study
MOZ99	L2a1a2	SA	Mozambique	-	Present study
SOM72	L2a1a2	EA	Somalia	-	Present study
DQ304975	L2a1a2a	AM	USA	-	[7]
JQ044968	L2a1a2a1	WA	Burkina Faso	-	[3]
JQ044845	L2a1a2a1	WA	Burkina Faso	-	[3]

Annex 2. *(continued)*

MOZ36	L2a1a2a1	SA	Mozambique	-	Present study
DQ304969	L2a1a2a1a	AM	USA	-	[7]
DQ304970	L2a1a2a1a	AM	USA	-	[7]
KJ185973	L2a1a2a1a	SA	Zambia	Kaonde	[9]
KJ185733	L2a1a2a1a	SA	Zambia	Nkoya	[9]
KJ185732	L2a1a2a1a	SA	Zambia	Nkoya	[9]
DQ304976	L2a1a2a1a	AM	USA	-	[7]
DQ304974	L2a1a2a1a	AM	USA	-	[7]
Howell142	L2a1a2a1a	AM	USA	-	[4]
EU092829	L2a1a2a1a	NA	Tunisia	-	[1]
JQ045017	L2a1a2a1a	WA	Burkina Faso	-	[3]
JQ705055	L2a1a2a1a	SA	Mozambique	-	[8]
KJ185826	L2a1a2a1a	SA	Angola	Ovimbundu	[9]
KJ185679	L2a1a2a1a	SA	Zambia	Mbunda	[9]
KJ185824	L2a1a2a1a	SA	Angola	Ovimbundu	[9]
MOZ317	L2a1a2a1a	SA	Mozambique	-	Present study
EU092872	L2a1a2a1a	SA	S. Africa	SEB	[1]
KJ185997	L2a1a2a1a	SA	Zambia	Ngoni	[9]
Howell562	L2a1a2a1a	AM	USA	-	[4]
HG02941	L2a1a2a1a	CA	Nigeria	Esan	[2]
EU092778	L2a1a2a1a	NA	Egypt	-	[1]
KJ185455	L2a1a2a1a	SA	Zambia	Lenje	[9]
DQ304972	L2a1a2b	AM	USA	-	[7]
DQ304973	L2a1a2b	AM	USA	-	[7]
Howell199	L2a1a2b	AM	USA	-	[4]
HG02439	L2a1a2b	AM	Barbados	African Caribbean	[2]
JQ045088	L2a1a2b	WA	Yoruba	Yoruba	[3]
JQ045095	L2a1a2b	WA	Yoruba	Yoruba	[3]
HG02676	L2a1a3	WA	Gambia	-	[2]
HG03539	L2a1a3	WA	Gambia	-	[2]
HG02870	L2a1a3	WA	Gambia	-	[2]
JQ044892	L2a1a3	WA	Burkina Faso	-	[3]
JQ045015	L2a1a3	WA	Burkina Faso	-	[3]
JQ044877	L2a1a3	WA	Burkina Faso	-	[3]
EU092711	L2a1a3	EUR	Portugal	Portuguese	[1]
EU092890	L2a1a3a	CA	Chad	Sara	[1]
EU092905	L2a1a3a	CA	Chad	Sara	[1]
JQ044813	L2a1a3a	WA	Burkina Faso	-	[3]
JQ705529	L2a1a3a	-	unknown	-	[8]
JQ044918	L2a1a3b	WA	Burkina Faso	-	[3]
JQ044819	L2a1a3b	WA	Burkina Faso	-	[3]
DQ304948	L2a1a3b	AM	USA	-	[7]
HM771168	L2a1a3c	CA	Pygmy	Pygmy	[14]
HG03169	L2a1a3c	CA	Nigeria	Esan	[2]
KJ185771	L2a1a3c	SA	Angola	Nyaneka	[9]

Annex 2. *(continued)*

KC622184	L2a1a3c	SA	Namibia	Himba	[10]
STP99	L2a1a3c	WA	STP	-	Present study
KJ185770	L2a1a3c	SA	Angola	Nyaneka	[9]
JQ045101	L2a1a3c	WA	Yoruba	Yoruba	[3]
HG02977	L2a1a3c	CA	Nigeria	Esan	[2]
JQ044975	L2a1b	WA	Burkina Faso	-	[3]
JQ044841	L2a1b	WA	Burkina Faso	-	[3]
NA20274	L2a1b	AM	USA	-	[2]
KJ185828	L2a1b	SA	Angola	Ovimbundu	[9]
EU092761	L2a1b	NA	Egypt	-	[1]
MOZ95	L2a1b1	SA	Mozambique	-	Present study
HG03132	L2a1b1	CA	Nigeria	Esan	[2]
HG02577	L2a1b1	AM	Barbados	African Caribbean	[2]
HG03202	L2a1b1	CA	Nigeria	Esan	[2]
MOZ64	L2a1b1a	SA	Mozambique	-	Present study
JX303858	L2a1b1a	SA	Zambia	Mbukushu	[13]
JX303761	L2a1b1a	SA	Zambia	Tonga	[13]
KJ185735	L2a1b1a	SA	Zambia	Nkoya	[9]
KJ185931	L2a1b1a	SA	Zambia	Makoma	[9]
KJ185825	L2a1b1a	SA	Angola	Ovimbundu	[9]
EU092690	L2a1b1a	SA	Mozambique	Chopi	[1]
EU092910	L2a1b1a	EA	Kenya	-	[1]
KC622066	L2a1b1a	SA	Botswana	Kalanga	[10]
MOZ318	L2a1b1a	SA	Mozambique	-	Present study
KJ185442	L2a1b1a	SA	Zambia	Luvale	[9]
JX303754	L2a1b1a	SA	Zambia	Tonga	[13]
KC622071	L2a1b1a	SA	Botswana	Kalanga	[10]
KJ185923	L2a1b1a	SA	Zambia	Luyana	[9]
JX303874	L2a1b1a	SA	Zambia	Tonga	[13]
JX303880	L2a1b1a	SA	Zambia	Totela	[13]
MOZ62	L2a1b1a	SA	Mozambique	-	Present study
MOZ45	L2a1b1a	SA	Mozambique	-	Present study
SOM13	L2a1b1a	EA	Somalia	-	Present study
SOM144	L2a1b1a	EA	Somalia	-	Present study
EU092705	L2a1b1a	SA	Mozambique	-	[1]
KC622057	L2a1b1a	SA	Botswana	Tswana	[10]
KJ185588	L2a1b1a	SA	Zambia	Lozi	[9]
KJ185452	L2a1b1a	SA	Zambia	Ila	[9]
KC622062	L2a1b1a	SA	Botswana	Tswana	[10]
KJ185587	L2a1b1a	SA	Zambia	Lozi	[9]
EU092919	L2a1b1a	AP / NE	Jordan	-	[1]
EF177417	L2a1c	EUR	Portugal	-	[15]
JQ702659	L2a1c	EUR	France	-	[8]
JN214436	L2a1c	EUR	Spain	Galicia	[11]
JQ044973	L2a1c	WA	Burkina Faso	-	[3]

Annex 2. *(continued)*

DQ304949	L2a1c	AM	USA	-	[7]
DQ304944	L2a1c	AM	USA	-	[7]
EU092901	L2a1c	CA	Chad	Sara	[1]
HM771224	L2a1c	CA	Pygmy	-	[14]
EU092733	L2a1c	WA	G. Bissau	Balanata	[1]
EU092937	L2a1c	EA	Ethiopia	-	[1]
JN858955	L2a1c	CA	Cameroon	-	[11]
JQ702968	L2a1c	-	unknown	-	[8]
EU092683	L2a1c	AP / NE	Israel	Palestinian	[1]
KJ185486	L2a1c	SA	Angola	Ganguela	[9]
JQ045019	L2a1c	WA	Burkina Faso	-	[3]
JQ702307	L2a1c	-	unknown	-	[8]
HG02808	L2a1c	WA	Gambia	-	[2]
JQ702503	L2a1c	AM	Mexico	-	[8]
HG02502	L2a1c	AM	Barbados	African Caribbean	[2]
HG02679	L2a1c	WA	Gambia	-	[2]
HG03078	L2a1c	WA	Sierra Leone	Mende	[2]
KJ185999	L2a1c	SA	Zambia	Nsenga	[9]
ETH48	L2a1c+16129	EA	Ethiopia	-	Present study
FJ460560	L2a1c1	NA	Tunisia	-	[5]
JQ044909	L2a1c1	WA	Burkina Faso	-	[3]
Howell561	L2a1c1	AM	USA	-	[4]
JQ044860	L2a1c1	WA	Burkina Faso	-	[3]
JQ044974	L2a1c1	WA	Burkina Faso	-	[3]
JQ045124	L2a1c1	WA	Mandenka	Mandenka	[3]
JX303852	L2a1c1	SA	Zambia	Subiya	[13]
KJ185731	L2a1c1	SA	Zambia	Nkoya	[9]
KJ185595	L2a1c1	SA	Zambia	Lozi	[9]
JQ045064	L2a1c1	WA	Burkina Faso	-	[3]
JQ705046	L2a1c1a	WA	Ghana	-	[8]
JQ044799	L2a1c1a	WA	Burkina Faso	-	[3]
Howell165	L2a1c1a	AM	USA	-	[4]
JQ702261	L2a1c1a	-	unknown	-	[8]
EU092954	L2a1c1a	EA	Ethiopia	-	[1]
JX524225	L2a1c1a	AM	Brazil	-	Family Tree
HG03117	L2a1c1a	CA	Nigeria	Esan	[2]
JQ044962	L2a1c2	WA	Burkina Faso	-	[3]
JQ045108	L2a1c2	WA	Mandenka	Mandenka	[3]
JQ044935	L2a1c2	WA	Burkina Faso	-	[3]
JQ044844	L2a1c2	WA	Burkina Faso	-	[3]
JQ044802	L2a1c2	WA	Burkina Faso	-	[3]
HG02819	L2a1c2	WA	Gambia	-	[2]
JQ044998	L2a1c2	WA	Burkina Faso	-	[3]
JQ045024	L2a1c2	WA	Burkina Faso	-	[3]
JQ044828	L2a1c2a	WA	Burkina Faso	-	[3]

Annex 2. *(continued)*

JQ044924	L2a1c2a	WA	Burkina Faso	-	[3]
HM771169	L2a1c2a	CA	Pygmy	-	[14]
Tor65(#27)	L2a1c2a	AM	Dominica	-	[16]
JQ044872	L2a1c2a	WA	Burkina Faso	-	[3]
JQ044969	L2a1c2a	WA	Burkina Faso	-	[3]
JQ045025	L2a1c2a	WA	Burkina Faso	-	[3]
HG03449	L2a1c2a	WA	Sierra Leone	Mende	[2]
EU200762	L2a1c3	EUR	Slovenia	-	[17]
EU092663	L2a1c3	AP / NE	Israel	Bedouin	[1]
JN214440	L2a1c3	EUR	Spain	Galicia	[11]
HG03557	L2a1c3	WA	Sierra Leone	Mende	[2]
HG03571	L2a1c3	WA	Sierra Leone	Mende	[2]
JQ705145	L2a1c3	AM	USA	-	[8]
HG02861	L2a1c3	WA	Gambia	-	[2]
HG02895	L2a1c3	WA	Gambia	-	[2]
EU092720	L2a1c3	WA	G. Bissau	FulaForro	[1]
JQ044996	L2a1c3	WA	Burkina Faso	-	[3]
JQ045063	L2a1c3	WA	Burkina Faso	-	[3]
HG02851	L2a1c3	WA	Gambia	-	[2]
JQ045122	L2a1c3	WA	Mandenka	Mandenka	[3]
JQ045110	L2a1c3	WA	Mandenka	Mandenka	[3]
JQ045116	L2a1c3	WA	Mandenka	Mandenka	[3]
HG03088	L2a1c3	WA	Sierra Leone	Mende	[2]
JN214432	L2a1c4	EUR	Spain	Andalusia	[11]
DQ304942	L2a1c4a	AM	USA	-	[7]
DQ304943	L2a1c4a	AM	USA	-	[7]
JQ044833	L2a1c4a	WA	Burkina Faso	-	[3]
JQ044957	L2a1c4a	WA	Burkina Faso	-	[3]
JQ045014	L2a1c4a	WA	Burkina Faso	-	[3]
JQ045035	L2a1c4a	WA	Burkina Faso	-	[3]
JQ045045	L2a1c4a	WA	Burkina Faso	-	[3]
JQ045067	L2a1c4a	WA	Burkina Faso	-	[3]
JQ701926	L2a1c4a	-	unknown	-	[8]
FJ460549	L2a1c4a	NA	Tunisia	-	[5]
DQ304951	L2a1c4a1	AM	USA	-	[7]
DQ304950	L2a1c4a1	AM	USA	-	[7]
KJ185487	L2a1c4a1	SA	Angola	Ganguela	[9]
KJ185488	L2a1c4a1	SA	Angola	Ganguela	[9]
Howell576	L2a1c5	AM	USA	-	[4]
JQ705001	L2a1c5	-	unknown	-	[8]
HM771225	L2a1c5	CA	Pygmy	-	[14]
HG02546	L2a1c5	AM	Barbados	African Caribbean	[2]
JQ045082	L2a1c5	WA	Yoruba	Yoruba	[3]
JQ045098	L2a1c5	WA	Yoruba	Yoruba	[3]
EU935443	L2a1c5	NA	Egypt	el-Hayez oasis	[18]

Annex 2. *(continued)*

HG02455	L2a1c5	AM	Barbados	African Caribbean	[2]
HG02549	L2a1c5	AM	Barbados	African Caribbean	[2]
JX303749	L2a1c5	SA	Zambia	Totela	[13]
KJ185585	L2a1c5	SA	Zambia	Lozi	[9]
KJ185584	L2a1c5	SA	Zambia	Lozi	[9]
KJ185468	L2a1c5	SA	Zambia	Tonga	[9]
NA20903	L2a1d	AM	USA	Gujarati Indian	[2]
EU092939	L2a1d1	EA	Ethiopia	-	[1]
SOM01	L2a1d1	EA	Somalia	-	Present study
EU092765	L2a1d1	NA	Egypt	-	[1]
EU092927	L2a1d1	NA	Egypt	-	[1]
SUD54	L2a1d1	EA	Sudan	-	Present study
JX303806	L2a1d2	SA	Zambia	Kwamashi	[13]
JX303862	L2a1d2	SA	Zambia	Mbukushu	[13]
JX303873	L2a1d2	SA	Zambia	Tonga	[13]
KJ185686	L2a1d2	SA	Zambia	Mbunda	[9]
KJ185591	L2a1d2	SA	Zambia	Lozi	[9]
JX303795	L2a1d2	SA	Zambia	Totela	[13]
JX303825	L2a1d2	SA	Zambia	Mbukushu	[13]
JX303838	L2a1d2	SA	Zambia	Mbukushu	[13]
KC622074	L2a1d2	SA	Botswana	Tswana	[10]
KJ185459	L2a1d2	SA	Zambia	Tokaleya	[9]
KJ185998	L2a1d2	SA	Zambia	Nsenga	[9]
KJ185691	L2a1d2	SA	Zambia	Mbunda	[9]
KJ185428	L2a1d2	SA	Zambia	Luchazi	[9]
KJ185689	L2a1d2	SA	Zambia	Mbunda	[9]
KJ185590	L2a1d2	SA	Zambia	Lozi	[9]
KJ185898	L2a1d2	SA	Zambia	Kwangwa	[9]
KJ185586	L2a1d2	SA	Zambia	Lozi	[9]
KJ185955	L2a1d2	SA	Zambia	Nyengo	[9]
KJ185939	L2a1d2	SA	Zambia	Mbowe	[9]
KJ185901	L2a1d2	SA	Zambia	Kwangwa	[9]
MOZ342	L2a1d2	SA	Mozambique	-	Present study
KJ185403	L2a1d2	SA	Zambia	Bemba	[9]
JN858956	L2a1d2	WA	Benin	-	[11]
Howell389	L2a1e	AM	USA	-	[4]
DQ304945	L2a1e	AM	USA	-	[7]
Howell172	L2a1e1	AM	USA	-	[4]
JQ705455	L2a1e1	-	unknown	-	[8]
JQ044821	L2a1e1	WA	Burkina Faso	-	[3]
JQ044970	L2a1e1	WA	Burkina Faso	-	[3]
DQ304946	L2a1e1	AM	USA	-	[7]
DQ304947	L2a1e1	AM	USA	-	[7]
Howell421	L2a1e1	AM	USA	-	[4]
DQ304930	L2a1e1	AM	USA	-	[7]

Annex 2. *(continued)*

DQ304929	L2a1e1	AM	USA	-	[7]
DQ304931	L2a1e1	AM	USA	-	[7]
JQ044883	L2a1f	WA	Burkina Faso	-	[3]
DQ304935	L2a1f	AM	USA	-	[7]
Howell566	L2a1f	AM	USA	-	[4]
KJ185678	L2a1f	SA	Zambia	Mbunda	[9]
Howell223	L2a1f	AM	USA	-	[4]
DQ304936	L2a1f	AM	USA	-	[7]
JQ044912	L2a1f	WA	Burkina Faso	-	[3]
JQ045056	L2a1f	WA	Burkina Faso	-	[3]
JQ045087	L2a1f	WA	Yoruba	Yoruba	[3]
HG03452	L2a1f	WA	Sierra Leone	Mende	[2]
KJ185728	L2a1f	SA	Zambia	Nkoya	[9]
KJ185899	L2a1f	SA	Zambia	Kwangwa	[9]
NA19023	L2a1f	EA	Kenya	Luhya	[2]
KJ185429	L2a1f	SA	Zambia	Luchazi	[9]
KJ185897	L2a1f	SA	Zambia	Kwangwa	[9]
KJ185893	L2a1f	SA	Zambia	Kwangwa	[9]
KJ185680	L2a1f	SA	Zambia	Mbunda	[9]
JX303832	L2a1f	SA	Zambia	Subiya	[13]
DQ304954	L2a1f	AM	USA	-	[7]
JQ705150	L2a1f	-	unknown	-	[8]
DQ304953	L2a1f	AM	USA	-	[7]
DQ304956	L2a1f	AM	USA	-	[7]
DQ304952	L2a1f	AM	USA	-	[7]
JQ704668	L2a1f	-	unknown	-	[8]
DQ304966	L2a1f	AM	USA	-	[7]
JQ703960	L2a1f	-	unknown	-	[8]
JQ044859	L2a1f	WA	Burkina Faso	-	[3]
Howell567	L2a1f	AM	USA	-	[4]
HG03159	L2a1f	CA	Nigeria	Esan	[2]
JQ045006	L2a1f	WA	Burkina Faso	-	[3]
HG03514	L2a1f	CA	Nigeria	Esan	[2]
JQ045061	L2a1f	WA	Burkina Faso	-	[3]
HG02979	L2a1f	CA	Nigeria	Esan	[2]
HG02885	L2a1f	WA	Gambia	-	[2]
HG03199	L2a1f	CA	Nigeria	Esan	[2]
AY195776	L2a1f	SA	S. Africa	-	[19]
KC622076	L2a1f	SA	Botswana	Tswana	[10]
KJ185681	L2a1f	SA	Zambia	Mbunda	[9]
KJ185683	L2a1f	SA	Zambia	Mbunda	[9]
KJ185985	L2a1f	SA	Zambia	Mambwe	[9]
Howell233	L2a1f	AM	USA	-	[4]
Howell565	L2a1f	AM	USA	-	[4]
DQ304959	L2a1f	AM	USA	-	[7]

Annex 2. *(continued)*

DQ304962	L2a1f	AM	USA	-	[7]
DQ304965	L2a1f	AM	USA	-	[7]
JQ045090	L2a1f	WA	Yoruba	Yoruba	[3]
DQ304960	L2a1f1	AM	USA	-	[7]
DQ304963	L2a1f1	AM	USA	-	[7]
DQ304964	L2a1f1	AM	USA	-	[7]
EU092961	L2a1f1	AM	USA	-	[1]
HG03575	L2a1f1	WA	Sierra Leone	Mende	[2]
JQ701814	L2a1f1	AM	USA	-	[8]
DQ304961	L2a1f1	AM	USA	-	[7]
DQ304955	L2a1f1	AM	USA	-	[7]
KJ185685	L2a1f1	SA	Zambia	Mbunda	[9]
HG02557	L2a1f1	AM	Barbados	African Caribbean	[2]
HG03135	L2a1f1	CA	Nigeria	Esan	[2]
KJ186002	L2a1f1	SA	Zambia	Tumbuka	[9]
KJ185933	L2a1f1	SA	Zambia	Makoma	[9]
DQ304967	L2a1f1	AM	USA	-	[7]
HG02476	L2a1f1	AM	Barbados	African Caribbean	[2]
HG03297	L2a1f1	CA	Nigeria	Esan	[2]
DQ304937	L2a1f1	AM	USA	-	[7]
HG03163	L2a1f1	CA	Nigeria	Esan	[2]
Howell208	L2a1f1	AM	USA	-	[4]
JX303780	L2a1f1	SA	Zambia	Tonga	[13]
DQ304958	L2a1f1a	AM	USA	-	[7]
DQ304934	L2a1f1a	AM	USA	-	[7]
JQ045097	L2a1f2	WA	Yoruba	Yoruba	[3]
JQ045039	L2a1f2	WA	Burkina Faso	-	[3]
DQ304957	L2a1f2	AM	USA	-	[7]
HG02481	L2a1f2	AM	Barbados	African Caribbean	[2]
EU597561	L2a1f3	EA	Kenya	-	[6]
JX303805	L2a1f3	SA	Zambia	Kwamashi	[13]
KJ185583	L2a1f3	SA	Zambia	Lozi	[9]
KJ185596	L2a1f3	SA	Zambia	Lozi	[9]
KJ185582	L2a1f3	SA	Zambia	Lozi	[9]
KC622118	L2a1f3	SA	Botswana	Kgalagadi	[10]
KJ185895	L2a1f3	SA	Zambia	Kwangwa	[9]
JQ044861	L2a1f3	WA	Burkina Faso	-	[3]
JX303752	L2a1g	SA	Zambia	Tonga	[13]
JX303906	L2a1g	SA	Zambia	Subiya	[13]
KJ185894	L2a1g	SA	Zambia	Kwangwa	[9]
KJ185693	L2a1g	SA	Zambia	Mbunda	[9]
JX303798	L2a1g	SA	Zambia	Totela	[13]
EU092676	L2a1h	AP / NE	Israel	Palestinian	[1]
EU092914	L2a1h	EA	Kenya	-	[1]
KJ186003	L2a1h	SA	Zambia	Tumbuka	[9]

Annex 2. *(continued)*

MOZ301	L2a1h	SA	Mozambique	-	Present study
JQ044905	L2a1i	WA	Burkina Faso	-	[3]
JQ045077	L2a1i	WA	Burkina Faso	-	[3]
Howell388	L2a1i	AM	USA	-	[4]
HG03432	L2a1i	WA	Sierra Leone	Mende	[2]
HG03461	L2a1i	WA	Sierra Leone	Mende	[2]
EU092719	L2a1i	WA	G. Bissau	FulaForro	Behar, 2008
HG03064	L2a1i	WA	Sierra Leone	Mende	[2]
JQ044958	L2a1i	WA	Burkina Faso	-	[3]
JQ045080	L2a1i	WA	Burkina Faso	-	[3]
JQ045102	L2a1i	WA	Yoruba	Yoruba	[3]
JQ044881	L2a1i1	WA	Burkina Faso	-	[3]
Howell193	L2a1i1	AM	USA	-	[4]
HG02554	L2a1i1	AM	Barbados	African Caribbean	[2]
AF346976	L2a1i1	-	unknown	Effik	[12]
NA18878	L2a1i1	CA	Nigeria	Yoruba	[2]
JX303857	L2a1i1	SA	Zambia	Fwe	[13]
KC622169	L2a1i1	SA	Namibia	Mbukushu	[10]
KJ185597	L2a1i1	SA	Zambia	Lozi	[9]
KJ185598	L2a1i1	SA	Zambia	Lozi	[9]
JX303909	L2a1i1	SA	Zambia	Subiya	[13]
KJ185690	L2a1i1	SA	Zambia	Mbunda	[9]
JX303853	L2a1i1	SA	Zambia	Fwe	[13]
KJ185900	L2a1i1	SA	Zambia	Kwangwa	[9]
SOM64	L2a1j	EA	Somalia	-	Present study
EU092756	L2a1j	AP / NE	Jordan	-	[1]
EU092816	L2a1j	NA	Morocco	Arab	[1]
EU200760	L2a1k	EUR	Czech Rep	-	[17]
EU200763	L2a1k	EUR	Slovenia	-	[17]
JQ045047	L2a1l	WA	Burkina Faso	-	[3]
JQ044956	L2a1l1	WA	Burkina Faso	-	[3]
JQ044919	L2a1l1	WA	Burkina Faso	-	[3]
JQ044955	L2a1l1	WA	Burkina Faso	-	[3]
HG01286	L2a1l1	AM	Puerto Rico	-	[2]
HG03428	L2a1l1	WA	Sierra Leone	Mende	[2]
HG03457	L2a1l1	WA	Sierra Leone	Mende	[2]
HG03382	L2a1l1	WA	Sierra Leone	Mende	[2]
HG03097	L2a1l1	WA	Sierra Leone	Mende	[2]
HG02568	L2a1l1	WA	Gambia	-	[2]
HG03401	L2a1l1	WA	Sierra Leone	Mende	[2]
HG03085	L2a1l1	WA	Sierra Leone	Mende	[2]
HG03547	L2a1l1	WA	Sierra Leone	Mende	[2]
JQ044817	L2a1l1a	WA	Burkina Faso	-	[3]
EU092812	L2a1l1a	NA	Morocco	Arab	[1]
FJ769771	L2a1l1a	AM	Bahamas	-	Family Tree

Annex 2. *(continued)*

JQ044966	L2a1l1a	WA	Burkina Faso	-	[3]
JQ044897	L2a1l1a	WA	Burkina Faso	-	[3]
DQ304939	L2a1l1a	AM	USA	-	[7]
JQ044932	L2a1l1a	WA	Burkina Faso	-	[3]
EU092807	L2a1l1a	NA	Morocco	-	[1]
EU092721	L2a1l2	WA	G. Bissau	Mandinga	[1]
HG02716	L2a1l2	WA	Gambia	-	[2]
HG02620	L2a1l2	WA	Gambia	-	[2]
HG02629	L2a1l2	WA	Gambia	-	[2]
JQ044978	L2a1l2	WA	Burkina Faso	-	[3]
JQ044994	L2a1l2	WA	Burkina Faso	-	[3]
JQ705185	L2a1l2a	EUR	Russia	Jew	[8]
JQ705049	L2a1l2a	EUR	Poland	Jew	[8]
Howell401	L2a1l2a	AM	USA	-	[4]
Howell564	L2a1l2a	AM	USA	-	[4]
EU092687	L2a1l2a	AP / NE	Israel	Ashkenazi	[1]
EU547188	L2a1l2a	EUR	Poland	Jew	Family Tree
JN204423	L2a1l2a	EUR	Poland	Jew	Family Tree
JX266264	L2a1l2a	EUR	Poland	Podhale	[20]
JQ705589	L2a1l2a1	EUR	Romania	-	[8]
EU564850	L2a1l2a2	-	unknown	-	Family Tree
JQ702015	L2a1l2a3	EUR	Poland	Jew	[8]
JQ702904	L2a1l2a4	-	unknown	-	[8]
JX266265	L2a1l2a5	EUR	Poland	Kashubia	[20]
JQ044804	L2a1m	WA	Burkina Faso	-	[3]
JQ045040	L2a1m	WA	Burkina Faso	-	[3]
JQ044916	L2a1m	WA	Burkina Faso	-	[3]
DQ304940	L2a1m1	AM	USA	-	[7]
Howell563	L2a1m1	AM	USA	-	[4]
JQ701914	L2a1m1	EUR	Ireland	-	[8]
DQ304938	L2a1m1	AM	USA	-	[7]
JQ044908	L2a1m1	WA	Burkina Faso	-	[3]
JQ703065	L2a1n	-	unknown	-	[8]
JQ044944	L2a1n	WA	Burkina Faso	-	[3]
DQ304941	L2a1n	AM	USA	-	[7]
JQ045000	L2a1o	WA	Burkina Faso	-	[3]
EU092739	L2a1o	AP / NE	Syria	-	[1]
JX303870	L2a1q	SA	Zambia	Fwe	[13]
KC622159	L2a1q	SA	Namibia	Kwanyama	[10]
NA19381	L2a1q	EA	Kenya	Luhya	[2]
EU092896	L2a2a	CA	Chad	Sara	[1]
SUD102	L2a2a	EA	Sudan	-	Present study
HM771205	L2a2a1	CA	Pygmy	-	[14]
SUD75	L2a2a1	EA	Sudan	-	Present study
EU092902	L2a2a1	CA	Chad	Sara	[1]

Annex 2. *(continued)*

EU092882	L2a2a1	CA	Chad	Sara	[1]
HM771191	L2a2b	CA	Pygmy	-	[14]
HM771207	L2a2b	CA	Pygmy	-	[14]
KJ185986	L2a2b	SA	Zambia	Mambwe	[9]
EU597525	L2a2b1	CA	Mbuti Pygmy	-	[6]
HM771193	L2a2b1	CA	Pygmy	-	[14]
AY195788	L2a2b1	SA	S. Africa	San	[19]
HM771192	L2a2b1	CA	Pygmy	-	[14]
EU597549	L2a2b1	CA	Mbuti Pygmy	-	[6]
HM771194	L2a2b1	CA	Pygmy	-	[14]
HM771208	L2a2b1	CA	Pygmy	-	[14]
HM771195	L2a2b1	CA	Pygmy	-	[14]
HM771206	L2a3	CA	Pygmy	-	[14]
HM771197	L2a4a	CA	Pygmy	-	[14]
HM771213	L2a4a	CA	Pygmy	-	[14]
HM771210	L2a4a	CA	Pygmy	-	[14]
HM771209	L2a4a	CA	Pygmy	-	[14]
HM771214	L2a4a	CA	Pygmy	-	[14]
HM771211	L2a4a	CA	Pygmy	-	[14]
HM771196	L2a4a	CA	Pygmy	-	[14]
HM771212	L2a4a	CA	Pygmy	-	[14]
HM771215	L2a4a	CA	Pygmy	-	[14]
HM596745	L2a5	AM	Bermuda	-	Family Tree
JX303829	L2a5	SA	Zambia	Kwamashi	[13]
KJ185684	L2a5	SA	Zambia	Mbunda	[9]
KJ185687	L2a5	SA	Zambia	Mbunda	[9]
KJ185682	L2a5	SA	Zambia	Mbunda	[9]
KJ185952	L2a5	SA	Zambia	Nyengo	[9]
KJ185734	L2a5	SA	Zambia	Nkoya	[9]
KJ185953	L2a5	SA	Zambia	Nyengo	[9]
KJ185592	L2a5	SA	Zambia	Lozi	[9]
KJ185954	L2a5	SA	Zambia	Nyengo	[9]
KJ185942	L2a5	SA	Zambia	Mwenyi	[9]
KJ185932	L2a5	SA	Zambia	Makoma	[9]
KJ185730	L2a5	SA	Zambia	Nkoya	[9]
KJ185729	L2a5	SA	Zambia	Nkoya	[9]
KJ185427	L2a5	SA	Zambia	Luchazi	[9]
KJ185830	L2a5	SA	Angola	Ovimbundu	[9]
KJ185441	L2a5	SA	Zambia	Luvale	[9]
KJ185593	L2a5	SA	Zambia	Lozi	[9]
KJ185594	L2a5	SA	Zambia	Lozi	[9]
KJ185823	L2a5	SA	Angola	Ovimbundu	[9]
NA19045	L2a5	EA	Kenya	Luhya	[2]
HQ384199	L2a5	EUR	Spain	-	[21]
KJ185525	L2a5	SA	Angola	Kuvale	[9]

Annex 2. *(continued)*

JN214443	L2b	EUR	Spain	Galicia	[11]
EU092734	L2b	WA	G. Bissau	Mandinga	[1]
HG02645	L2b	WA	Gambia	-	[2]
JQ045008	L2b	WA	Burkina Faso	-	[3]
EU092747	L2b1	AP / NE	S. Arabia	-	[1]
EU092766	L2b1	NA	Egypt	-	[1]
HG02982	L2b1a	WA	Gambia	-	[2]
JQ044854	L2b1a	WA	Burkina Faso	-	[3]
HG03291	L2b1a	CA	Nigeria	Esan	[2]
Tor66(#26)	L2b1a	AM	Dominica	-	[16]
HG01403	L2b1a	AM	Puerto Rico	-	[2]
DQ304981	L2b1a	AM	USA	-	[7]
HG02760	L2b1a	WA	Gambia	-	[2]
Howell175	L2b1a	AM	USA	-	[4]
HG03472	L2b1a	WA	Sierra Leone	Mende	[2]
EU092664	L2b1a2	AP / NE	Israel	Bedouin	[1]
EU092722	L2b1a2	WA	G. Bissau	Manjaco	[1]
DQ304985	L2b1a2	AM	USA	-	[7]
HG02887	L2b1a2	WA	Gambia	-	[2]
HG03046	L2b1a2	WA	Gambia	-	[2]
JQ045013	L2b1a2	WA	Burkina Faso	-	[3]
JQ044890	L2b1a2	WA	Burkina Faso	-	[3]
JQ045037	L2b1a2	WA	Burkina Faso	-	[3]
DQ304978	L2b1a3	AM	USA	-	[7]
DQ304979	L2b1a3	AM	USA	-	[7]
DQ304980	L2b1a3	AM	USA	-	[7]
DQ304982	L2b1a3	AM	USA	-	[7]
DQ304983	L2b1a3	AM	USA	-	[7]
DQ304984	L2b1a3	AM	USA	-	[7]
HG03515	L2b1a3	CA	Nigeria	Esan	[2]
AY195766	L2b1a3	SA	S. Africa	-	[19]
EU092854	L2b1a3	SA	S. Africa	San	[1]
KJ185460	L2b1a3	SA	Zambia	Tokaleya	[9]
KJ185857	L2b1a3	SA	Zambia	Ovimbundu	[9]
FJ460535	L2b1a3	NA	Tunisia	-	[5]
KJ185956	L2b1a3	SA	Zambia	Nyengo	[9]
KJ185832	L2b1a3	SA	Angola	Ovimbundu	[9]
KJ185489	L2b1a3	SA	Angola	Ganguela	[9]
KJ185772	L2b1a3	SA	Angola	Nyaneka	[9]
KJ185443	L2b1a3	SA	Zambia	Luvale	[9]
KJ185600	L2b1a3	SA	Zambia	Lozi	[9]
JQ701833	L2b1a3	EUR	unknown	African European	[8]
JQ702694	L2b1a3	-	unknown	-	[8]
JX303882	L2b1a3	SA	Zambia	Shanjo	[13]
KJ185869	L2b1a3	SA	Zambia	Kwamulonga	[9]

Annex 2. *(continued)*

KJ185599	L2b1a3	SA	Zambia	Lozi	[9]
NA19024	L2b1a3	EA	Kenya	Luhya	[2]
Howell568	L2b1a4	AM	USA	-	[4]
JN214453	L2b1a4	EUR	Italy	Liguria	[11]
JN214454	L2b1a4	EUR	Italy	Liguria	[11]
JQ044800	L2b1b	WA	Burkina Faso	-	[3]
HG02952	L2b1b	CA	Nigeria	Esan	[2]
Howell574	L2b1b	AM	USA	-	[4]
FJ228403	L2b1b	WA	Senegal	-	Family Tree
HM771226	L2b1b	CA	Pygmy	-	[14]
KJ185444	L2b1b	SA	Zambia	Luvale	[9]
JQ044910	L2b2	WA	Burkina Faso	-	[3]
EU092692	L2b2	SA	Mozambique	Ronga	[1]
KJ185974	L2b2	SA	Zambia	Kaonde	[9]
SUD87	L2b2	EA	Sudan	-	Present study
JQ044797	L2b2	WA	Burkina Faso	-	[3]
JQ045043	L2b2	WA	Burkina Faso	-	[3]
JQ044846	L2b2a	WA	Burkina Faso	-	[3]
STP84	L2b2a	WA	STP	-	Present study
JX303841	L2b2a	SA	Zambia	Kwamashi	[13]
KJ185694	L2b2a	SA	Zambia	Mbunda	[9]
JX303807	L2b2a	SA	Zambia	Kwamashi	[13]
KJ185833	L2b2a	SA	Angola	Ovimbundu	[9]
KJ185831	L2b2a	SA	Angola	Ovimbundu	[9]
EU092661	L2b3	EA	Ethiopia	-	[1]
FJ460526	L2b3	NA	Tunisia	-	[5]
HG02804	L2b3	WA	Gambia	-	[2]
HG02595	L2b3	WA	Gambia	-	[2]
HG02837	L2b3	WA	Gambia	-	[2]
HG02555	L2b3	AM	Barbados	African Caribbean	[2]
JQ702123	L2b3a	AM	Hawaii	-	[8]
Howell222	L2b3a	AM	USA	-	[4]
JQ702626	L2b3a	-	unknown	-	[8]
Howell385	L2b3a	AM	USA	-	[4]
JQ044882	L2c	WA	Burkina Faso	-	[3]
STP43	L2c	WA	STP	-	Present study
JQ701954	L2c	-	unknown	-	[8]
HG03484	L2c	WA	Sierra Leone	Mende	[2]
EU092723	L2c	WA	G. Bissau	Fula	[1]
JQ705120	L2c	-	unknown	-	[8]
HG03376	L2c	WA	Sierra Leone	Mende	[2]
HG03451	L2c	WA	Sierra Leone	Mende	[2]
HG03473	L2c	WA	Sierra Leone	Mende	[2]
HG02703	L2c	WA	Gambia	-	[2]
JQ702169	L2c	-	unknown	-	[8]

Annex 2. *(continued)*

JQ044878	L2c	WA	Burkina Faso	-	[3]
JQ044941	L2c	WA	Burkina Faso	-	[3]
HG03049	L2c	WA	Gambia	-	[2]
KJ185773	L2c	SA	Angola	Nyaneka	[9]
HG03433	L2c	WA	Sierra Leone	Mende	[2]
HG03455	L2c	WA	Sierra Leone	Mende	[2]
HG03476	L2c	WA	Sierra Leone	Mende	[2]
HG03091	L2c	WA	Sierra Leone	Mende	[2]
HG03225	L2c	WA	Sierra Leone	Mende	[2]
HG02814	L2c	WA	Gambia	-	[2]
HG02882	L2c	WA	Gambia	-	[2]
HG02888	L2c	WA	Gambia	-	[2]
HG02624	L2c	WA	Gambia	-	[2]
HG02642	L2c	WA	Gambia	-	[2]
JQ045030	L2c	WA	Burkina Faso	-	[3]
HG03072	L2c	WA	Sierra Leone	Mende	[2]
HG02879	L2c	WA	Gambia	-	[2]
STP48	L2c	WA	STP	-	Present study
JQ044989	L2c1	WA	Burkina Faso	-	[3]
JQ044858	L2c1	WA	Burkina Faso	-	[3]
JQ045002	L2c1	WA	Burkina Faso	-	[3]
HG02759	L2c1	WA	Gambia	-	[2]
HG02505	L2c1	AM	Barbados	African Caribbean	[2]
EU092813	L2c1a	NA	Morocco	-	[1]
JN214433	L2c1a	EUR	Spain	Andalusia	[11]
HG03048	L2c1a	WA	Gambia	-	[2]
JQ045106	L2c1a	WA	Mandenka	Mandenka	[3]
JQ045105	L2c1a	WA	Mandenka	Mandenka	[3]
JQ045068	L2c1a	WA	Burkina Faso	-	[3]
JQ045022	L2c1a	WA	Burkina Faso	-	[3]
JQ044887	L2c1a	WA	Burkina Faso	-	[3]
JQ044901	L2c1a	WA	Burkina Faso	-	[3]
AF381981	L2c1a	NA	Mauritania	-	[22]
EU092754	L2c2	AP / NE	Lebanon	-	[1]
EU092697	L2c2	SA	Mozambique	Mozambique	[1]
JQ045042	L2c2	WA	Burkina Faso	-	[3]
Tor67(#08)	L2c2	AM	Dominica	-	[16]
DQ304986	L2c2	AM	USA	-	[7]
Howell573	L2c2	AM	USA	-	[4]
HG03069	L2c2	WA	Sierra Leone	Mende	[2]
HG03388	L2c2	WA	Sierra Leone	Mende	[2]
DQ304989	L2c2	AM	USA	-	[7]
JQ044920	L2c2	WA	Burkina Faso	-	[3]
JQ045010	L2c2	WA	Burkina Faso	-	[3]
HG02836	L2c2	WA	Gambia	-	[2]

Annex 2. *(continued)*

DQ304988	L2c2a	AM	USA	-	[7]
EU092955	L2c2a	AM	USA	-	[1]
EU092957	L2c2a	AM	USA	-	[1]
JQ704740	L2c2a1	-	unknown	-	[8]
HG03157	L2c2a1	CA	Nigeria	Esan	[2]
DQ304987	L2c2a1	AM	USA	-	[7]
KJ185994	L2c2a1	SA	Zambia	Ndundulu	[9]
KJ185603	L2c2a1	SA	Zambia	Lozi	[9]
KJ185877	L2c2a1	SA	Zambia	Kwandi	[9]
KJ185601	L2c2a1	SA	Zambia	Lozi	[9]
KJ185605	L2c2a1	SA	Zambia	Lozi	[9]
KJ185539	L2c2a1	SA	Zambia	Kololo	[9]
KJ185983	L2c2a1	SA	Zambia	Lunda	[9]
JX303792	L2c2a1	SA	Zambia	Totela	[13]
KJ185604	L2c2a1	SA	Zambia	Lozi	[9]
JX303863	L2c2a1	SA	Zambia	Fwe	[13]
JX303750	L2c2a1	SA	Zambia	Totela	[13]
HG03354	L2c2b1	CA	Nigeria	Esan	[2]
JQ704094	L2c2b1a	-	unknown	-	[8]
NA18517	L2c2b1a	CA	Nigeria	Yoruba	[2]
EU092710	L2c2b1b	EUR	Netherlands	Dutch	[1]
KJ185834	L2c2b1b	SA	Angola	Ovimbundu	[9]
KC622075	L2c2b1b	SA	Botswana	Kalanga	[10]
KJ185775	L2c2b1b	SA	Angola	Nyaneka	[9]
KJ185526	L2c2b1b	SA	Angola	Kuvale	[9]
KJ185774	L2c2b1b	SA	Angola	Nyaneka	[9]
KJ185602	L2c2b1b	SA	Angola	Lozi	[9]
JQ044853	L2c2b2	WA	Burkina Faso	-	[3]
JQ044917	L2c3	WA	Burkina Faso	-	[3]
HG03074	L2c3	WA	Sierra Leone	Mende	[2]
JQ044810	L2c3	WA	Burkina Faso	-	[3]
JQ044971	L2c3	WA	Burkina Faso	-	[3]
JQ705626	L2c3	-	unknown	-	[8]
Howell575	L2c3	AM	USA	-	[4]
Howell572	L2c3	AM	USA	-	[4]
AF346995	L2c3a	-	unknown	Mandenka	[12]
JQ045104	L2c3a	WA	Mandenka	Mandenka	[3]
JQ045109	L2c3a	WA	Mandenka	Mandenka	[3]
JQ702115	L2c4	-	unknown	-	[8]
HG02757	L2c4	WA	Gambia	-	[2]
HG02878	L2c4	WA	Gambia	-	[2]
HG02756	L2c4	WA	Gambia	-	[2]
JQ044914	L2c4	WA	Burkina Faso	-	[3]
JQ044921	L2c4	WA	Burkina Faso	-	[3]
HG02715	L2c5	WA	Gambia	-	[2]

Annex 2. *(continued)*

AY195785	L2c5	SA	S. Africa	S. Africa	[19]
HG02643	L2c5	WA	Gambia	-	[2]
HG02896	L2c5	WA	Gambia	-	[2]
HG02462	L2c5	WA	Gambia	-	[2]
HG02667	L2c5	WA	Gambia	-	[2]
EU597570	L2d	AM	unknown	Latin America	[6]
SUD86	L2d	EA	Sudan	-	Present study
HG02536	L2d	AM	Barbados	African Caribbean	[2]
HG02881	L2d	WA	Gambia	-	[2]
EU092817	L2d	NA	Algeria	Arab	[1]
HG02799	L2d	WA	Gambia	-	[2]
HG02545	L2d1	AM	Barbados	African Caribbean	[2]
HG03437	L2d1	WA	Sierra Leone	Mende	[2]
HG03279	L2d1	CA	Nigeria	Esan	[2]
JQ045050	L2d1a	WA	Burkina Faso	-	[3]
HG02970	L2d1a	CA	Nigeria	Esan	[2]
KJ185421	L2d1a	SA	Zambia	Chokwe	[9]
JQ044948	L2d1a	WA	Burkina Faso	-	[3]
JQ045044	L2d1a	WA	Burkina Faso	-	[3]
EU092794	L2d1a	AP / NE	Yemen	-	[1]
JQ045060	L2d1a	WA	Burkina Faso	-	[3]
Howell160	L2d1a	AM	USA	-	[4]
DQ341062	L2d1a	EA	Ethiopia	-	[23]
JQ044929	L2d1a	WA	Burkina Faso	-	[3]
JQ045011	L2d1a	WA	Burkina Faso	-	[3]
JX303748	L2d1a	SA	Zambia	Totela	[13]
Tor64(#28)	L2e	AM	Dominica	-	[16]
HG03209	L2e	WA	Sierra Leone	Mende	[2]
HG03240	L2e	WA	Gambia	-	[2]
EU092724	L2e	WA	G. Bissau	Mandinga	[1]
SUD103	L2e1	EA	Sudan	-	Present study
JQ044816	L2e1	WA	Burkina Faso	-	[3]
Howell153	L2e1	AM	USA	-	[4]
STP12	L2e1	WA	STP	-	Present study
FJ460523	L2e1a	NA	Tunisia	-	[5]
HG03367	L2e1a	CA	Nigeria	Esan	[2]
HG03130	L2e1a	CA	Nigeria	Esan	[2]
KJ185902	L2e1a	SA	Zambia	Kwangwa	[9]
KJ185607	L2e1a	SA	Zambia	Lozi	[9]
KJ185695	L2e1a	SA	Zambia	Mbunda	[9]
KJ185608	L2e1a	SA	Zambia	Lozi	[9]
KJ185606	L2e1a	SA	Zambia	Lozi	[9]

References (Annex 2)

1. Behar DM, Villems R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, Scozzari R, Makkan H, Tzur S, Comas D, Bertranpetit J, Quintana-Murci L, Tyler-Smith C, Wells RS, Rosset S: **The Dawn of Human Matrilineal Diversity**. *The American Journal of Human Genetics* 2008, **82**:1130–1140.
2. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA: **An integrated map of genetic variation from 1,092 human genomes.** *Nature* 2012, **491**:56–65.
3. Barbieri C, Whitten M, Beyer K, Schreiber H, Li M, Pakendorf B: **Contrasting maternal and paternal histories in the linguistic context of Burkina Faso.** *Molecular Biology and Evolution* 2012, **29**:1213–23.
4. Howell N, Elson JL, Turnbull DM, Herrnstadt C: **African Haplogroup L mtDNA sequences show violations of clock-lfike evolution.** *Molecular Biology and Evolution* 2004, **21**:1843–54.
5. Costa MD, Cherni L, Fernandes V, Freitas F, Ammar El Gaaied A Ben, Pereira L: **Data from complete mtDNA sequencing of Tunisian centenarians: testing haplogroup association and the "golden mean" to longevity.** *Mechanisms of Ageing and Development* 2009, **130**:222–6.
6. Hartmann A, Thieme M, Nanduri LK, Stempfl T, Moehle C, Kivisild T, Oefner PJ: **Validation of microarray-based resequencing of 93 worldwide mitochondrial genomes.** *Human Mutation* 2009, **30**:115–22.
7. Just RS, Diegoli TM, Saunier JL, Irwin JA, Parsons TJ: **Complete mitochondrial genome sequences for 265 African American and U.S. "Hispanic" individuals.** *Forensic Science International Genetics* 2008, **2**:e45–8.
8. Behar DM, van Oven M, Rosset S, Metspalu M, Loogväli E-L, Silva NM, Kivisild T, Torroni A, Villems R: **A "Copernican" reassessment of the human mitochondrial DNA tree from its root.** *American Journal of Human Genetics* 2012, **90**:675–84.
9. Barbieri C, Vicente M, Oliveira S, Bostoen K, Rocha J, Stoneking M, Pakendorf B: **Migration and Interaction in a Contact Zone: mtDNA Variation among Bantu-Speakers in Southern Africa.** *PloS one* 2014, **9**:e99117.
10. Barbieri C, Güldemann T, Naumann C, Gerlach L, Berthold F, Nakagawa H, Mpoloka SW, Stoneking M, Pakendorf B: **Unraveling the complex maternal history of Southern African Khoisan populations.** *American Journal of Physical Anthropology* 2014, **153**:435–48.
11. Cerezo M, Achilli A, Olivieri A, Perego U a, Gómez-Carballa A, Brisighelli F, Lancioni H, Woodward SR, López-Soto M, Carracedo A, Capelli C, Torroni A, Salas A: **Reconstructing ancient mitochondrial DNA links between Africa and Europe.** *Genome Research* 2012, **22**:821–6.
12. Ingman M, Kaessmann H, Pääbo S, Gyllensten U: **Mitochondrial genome variation and the origin of modern humans.** *Nature* 2000, **408**:708–13.
13. Barbieri C, Butthof A, Bostoen K, Pakendorf B: **Genetic perspectives on the origin of clicks in Bantu languages from southwestern Zambia.** *European Journal of Human Genetics* 2013, **21**:430–6.
14. Batini C, Lopes J, Behar DM, Calafell F, Jorde LB, van der Veen L, Quintana-Murci L, Spedini G, Destro-Bisol G, Comas D: **Insights into the demographic history of African Pygmies from complete mitochondrial genomes.** *Molecular Biology and Evolution* 2011, **28**:1099–110.
15. Pereira L, Gonçalves J, Franco-Duarte R, Silva J, Rocha T, Arnold C, Richards M, Macaulay V: **No Evidence for an mtDNA Role in Sperm Motility: Data from Complete Sequencing of Asthenozoospermic Males**. *Molecular Biology and Evolution* 2007, **24** (3):868–874.
16. Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, Calderon FL, Simionati B, Valle G, Richards M, Macaulay V, Scozzari R: **Do the Four Clades of the mtDNA Haplogroup L2 Evolve at Different Rates?** *American Journal of Human Genetics* 2001, **69**:1348–1356.
17. Malyarchuk BA, Derenko M, Perkova M, Grzybowski T, Vanecek T, Lazur J: **Reconstructing the phylogeny of African mitochondrial DNA lineages in Slavs.** *European Journal of Human Genetics* 2008, **16**:1091–6.

18. Kujanová M, Pereira L, Fernandes V, Pereira JB, Cerný V: **Near eastern neolithic genetic input in a small oasis of the Egyptian Western Desert.** *American Journal of Physical Anthropology* 2009, **140**:336–46.
19. Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC: **Natural selection shaped regional mtDNA variation in humans.** *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**:171–6.
20. Mielnik-Sikorska M, Daca P, Malyarchuk B: **The History of Slavs Inferred from Complete Mitochondrial Genome Sequences**. *PloS one* 2013, **8**:e54360.
21. Gómez-Carballa A, Cerezo M, Balboa E, Heredia C, Castro-Feijóo L, Rica I, Barreiro J, Eirís J, Cabanas P, Martínez-Soto I, Fernández-Toral J, Castro-Gago M, Pombo M, Carracedo Á, Barros F, Salas A: **Evolutionary analyses of entire genomes do not support the association of mtDNA mutations with Ras/MAPK pathway syndromes.** *PloS one* 2011, **6**:e18348.
22. Maca-Meyer N, Gonzalez A, Larruga J, Flores C, Cabrera V: **Major genomic mitochondrial lineages delineate early human expansions**. *BMC Genetics* 2001, **2**:13.
23. Torroni A, Achilli A, Macaulay V, Richards M, Bandelt H-J: **Harvesting the fruit of the human mtDNA tree.** *Trends in Genetics* 2006, **22**:339–45.

Annex 3. List of 303 complete L0a sequences used for phylogenetic reconstruction and phylogeography analysis.

Codes for regions as in Table 1. Abbreviations: CAR – Central African Republic, S. Africa – South Africa, STP – São Tomé and Príncipe, S. Arabia – Saudi Arabia, USA – United States of America. Additional information on ethnicity or geographical location indicated whenever provided by the authors.

Sample	Haplogroup	Region	Country / Ethnic group	Additional information	Reference
NA19039	L0a+95	EA	Kenya	Luhya	[1]
JX303766	L0a1	SA	Zambia	Shanjo	[2]
KF672808	L0a1	SA	Mozambique	-	[3]
KJ186009	L0a1	SA	Angola	Ovimbundu	[4]
KJ185432	L0a1+16293	SA	Zambia	Luvale	[4]
EU092665	L0a1a	AP / NE	Israel	Bedouin	[5]
EU092764	L0a1a	NA	Egypt	-	[5]
KF672829	L0a1a	CA	Chad	Kanembou	[3]
EU092763	L0a1a+200	NA	Egypt	-	[5]
EU092881	L0a1a+200	CA	Chad	Laal	[5]
EU092892	L0a1a+200	CA	Chad	Sara	[5]
HG03063	L0a1a+200	WA	Sierra Leone	Mende	[1]
JQ044851	L0a1a+200	WA	Burkina Faso	-	[6]
JQ044903	L0a1a+200	WA	Burkina Faso	-	[6]
JQ702428	L0a1a+200	EUR	Italy	-	[7]
KF672821	L0a1a+200	EA	Ethiopia	-	[3]
KF672830	L0a1a+200	EA	Ethiopia	Oromo	[3]
NA19027	L0a1a+200	EA	Kenya	Luhya	[1]
NA19042	L0a1a+200	EA	Kenya	Luhya	[1]
NA19156	L0a1a+200	CA	Nigeria	Yoruba	[1]
NA19311	L0a1a+200	EA	Kenya	Luhya	[1]
NA19350	L0a1a+200	EA	Kenya	Luhya	[1]
NA19379	L0a1a+200	EA	Kenya	Luhya	[1]
EU092714	L0a1a1	WA	G. Bissau	Beafada	[5]
JQ044893	L0a1a1	WA	Burkina Faso	-	[6]
KF672836	L0a1a1	EA	Sudan	Arab	[3]
KF672837	L0a1a1	EA	Sudan	Nubian	[3]
AF346985	L0a1a2	-	unknown	Hausa	[8]
AY195780	L0a1a2	SA	S. Africa	-	[9]
DQ304901	L0a1a2	AM	USA	-	[10]
DQ304902	L0a1a2	AM	USA	-	[10]
DQ304903	L0a1a2	AM	USA	-	[10]
DQ304904	L0a1a2	AM	USA	-	[10]
EU092819	L0a1a2	NA	Algeria	Kabyle	[5]
HG01108	L0a1a2	AM	Puerto Rico	-	[1]
HG02971	L0a1a2	CA	Nigeria	Esan	[1]
HG02981	L0a1a2	CA	Nigeria	Esan	[1]
HG03127	L0a1a2	CA	Nigeria	Esan	[1]
HG03267	L0a1a2	CA	Nigeria	Esan	[1]

Annex 3. *(continued)*

HG03268	L0a1a2	CA	Nigeria	Esan	[1]
HG03298	L0a1a2	CA	Nigeria	Esan	[1]
HG03517	L0a1a2	CA	Nigeria	Esan	[1]
JQ044838	L0a1a2	WA	Burkina Faso	-	[6]
JQ044849	L0a1a2	WA	Burkina Faso	-	[6]
JQ044874	L0a1a2	WA	Burkina Faso	-	[6]
JQ044995	L0a1a2	WA	Burkina Faso	-	[6]
JQ045004	L0a1a2	WA	Burkina Faso	-	[6]
JQ045053	L0a1a2	WA	Burkina Faso	-	[6]
JQ702227	L0a1a2	-	unknown	-	[7]
JQ703481	L0a1a2	-	unknown	-	[7]
JX303757	L0a1a2	SA	Zambia	Subiya	[2]
JX303869	L0a1a2	SA	Zambia	Fwe	[2]
JX303911	L0a1a2	SA	Zambia	Tonga	[2]
KF672822	L0a1a2	WA	STP	-	[3]
KF672826	L0a1a2	WA	STP	-	[3]
KJ185430	L0a1a2	SA	Zambia	Luvale	[4]
KJ185656	L0a1a2	SA	Zambia	Mbunda	[4]
KJ185754	L0a1a2	SA	Angola	Nyaneka	[4]
KJ185806	L0a1a2	SA	Angola	Ovimbundu	[4]
KJ185863	L0a1a2	SA	Zambia	Tswana	[4]
KJ185884	L0a1a2	SA	Zambia	Kwangwa	[4]
KJ185972	L0a1a2	SA	Zambia	Kaonde	[4]
KJ186004	L0a1a2	SA	Zambia	Yeyi	[4]
KJ186005	L0a1a2	SA	Zambia	Yeyi	[4]
NA18861	L0a1a2	CA	Nigeria	Yoruba	[1]
NA18876	L0a1a2	CA	Nigeria	Yoruba	[1]
NA18877	L0a1a2	CA	Nigeria	Yoruba	[1]
NA19137	L0a1a2	CA	Nigeria	Yoruba	[1]
NA19216	L0a1a2	CA	Nigeria	Yoruba	[1]
NA19430	L0a1a2	EA	Kenya	Luhya	[1]
NA20334	L0a1a2	AM	USA	-	[1]
NA20336	L0a1a2	AM	USA	-	[1]
NA20355	L0a1a2	AM	USA	-	[1]
HG03084	L0a1a3	WA	Sierra Leone	Mende	[1]
Howell586	L0a1a3	AM	USA	-	[11]
JQ044943	L0a1a3	WA	Burkina Faso	-	[6]
NA18510	L0a1a3	CA	Nigeria	Yoruba	[1]
EU092878	L0a1b	CA	Chad	Laal	[5]
EU092889	L0a1b	CA	Chad	Sara	[5]
Howell585	L0a1b	AM	USA	-	[11]
AF381988	L0a1b1	NA	Morocco	-	[12]
DQ304900	L0a1b1	AM	USA	-	[10]
JX303762	L0a1b1	SA	Zambia	Fwe	[2]
KJ185398	L0a1b1	SA	Zambia	Bemba	[4]
KJ185510	L0a1b1	SA	Angola	Kuvale	[4]
DQ304899	L0a1b1a	AM	USA	-	[10]
EU935434	L0a1b1a	NA	Egypt	el-Hayez oasis	[13]

Annex 3. *(continued)*

EU935437	L0a1b1a	NA	Egypt	el-Hayez oasis	[13]
EU935464	L0a1b1a	NA	Egypt	el-Hayez oasis	[13]
EU935467	L0a1b1a	NA	Egypt	el-Hayez oasis	[13]
JX303817	L0a1b1a	SA	Zambia	Fwe	[2]
JX303823	L0a1b1a	SA	Zambia	Fwe	[2]
KF672807	L0a1b1a	CA	Chad	Daza	[3]
KF672827	L0a1b1a	CA	Cameroon	Kotoco	[3]
KJ185504	L0a1b1a	SA	Angola	Kuvale	[4]
KJ185509	L0a1b1a	SA	Angola	Kuvale	[4]
KJ185753	L0a1b1a	SA	Angola	Nyaneka	[4]
KJ185756	L0a1b1a	SA	Angola	Nyaneka	[4]
KJ185757	L0a1b1a	SA	Angola	Nyaneka	[4]
KJ185759	L0a1b1a	SA	Angola	Nyaneka	[4]
NA19703	L0a1b1a	AM	USA	-	[1]
EU092688	L0a1b1a1	SA	Mozambique	Shangaan	[5]
EU092858	L0a1b1a1	SA	S. Africa	San	[5]
EU092869	L0a1b1a1	SA	S. Africa	SWB	[5]
EU092871	L0a1b1a1	SA	S. Africa	SEB	[5]
EU092909	L0a1b1a1	EA	Kenya	-	[5]
Howell587	L0a1b1a1	AM	USA	-	[11]
JQ702326	L0a1b1a1	-	unknown	-	[7]
JX303796	L0a1b1a1	SA	Zambia	Kwamashi	[2]
KC622063	L0a1b1a1	SA	Botswana	Tswana	[14]
KC622067	L0a1b1a1	SA	Botswana	Tswana	[14]
KC622068	L0a1b1a1	SA	Botswana	Tswana	[14]
KC622077	L0a1b1a1	SA	Botswana	Kgalagadi	[14]
KC622078	L0a1b1a1	SA	Botswana	Kalanga	[14]
KC622112	L0a1b1a1	SA	Botswana	Kgalagadi	[14]
KC622114	L0a1b1a1	SA	Botswana	Kgalagadi	[14]
KC622115	L0a1b1a1	SA	Botswana	Kgalagadi	[14]
KC622117	L0a1b1a1	SA	Botswana	Kgalagadi	[14]
KC622121	L0a1b1a1	SA	Namibia	Owambo	[14]
KF672805	L0a1b1a1	SA	Mozambique	-	[3]
KF672806	L0a1b1a1	SA	Mozambique	-	[3]
KJ185423	L0a1b1a1	SA	Zambia	Luchazi	[4]
KJ185425	L0a1b1a1	SA	Zambia	Luchazi	[4]
KJ185431	L0a1b1a1	SA	Zambia	Luvale	[4]
KJ185540	L0a1b1a1	SA	Zambia	Lozi	[4]
KJ185542	L0a1b1a1	SA	Zambia	Lozi	[4]
KJ185545	L0a1b1a1	SA	Zambia	Lozi	[4]
KJ185547	L0a1b1a1	SA	Zambia	Lozi	[4]
KJ185548	L0a1b1a1	SA	Zambia	Lozi	[4]
KJ185549	L0a1b1a1	SA	Zambia	Lozi	[4]
KJ185552	L0a1b1a1	SA	Zambia	Lozi	[4]
KJ185554	L0a1b1a1	SA	Zambia	Lozi	[4]
KJ185555	L0a1b1a1	SA	Zambia	Lozi	[4]
KJ185659	L0a1b1a1	SA	Zambia	Mbunda	[4]
KJ185660	L0a1b1a1	SA	Zambia	Mbunda	[4]

Annex 3. *(continued)*

KJ185718	L0a1b1a1	SA	Zambia	Nkoya	[4]
KJ185858	L0a1b1a1	SA	Zambia	Shanjo	[4]
KJ185966	L0a1b1a1	SA	Zambia	Yauma	[4]
KJ185967	L0a1b1a1	SA	Zambia	Shona	[4]
KJ185977	L0a1b1a1	SA	Zambia	Lunda	[4]
KJ185978	L0a1b1a1	SA	Zambia	Lunda	[4]
KJ185979	L0a1b1a1	SA	Zambia	Lunda	[4]
KJ185990	L0a1b1a1	SA	Zambia	-	[4]
KJ186001	L0a1b1a1	SA	Zambia	Tebele	[4]
NA19382	L0a1b1a1	EA	Kenya	Luhya	[1]
NA19448	L0a1b1a1	EA	Kenya	Luhya	[1]
NA19449	L0a1b1a1	EA	Kenya	Luhya	[1]
NA19466	L0a1b1a1	EA	Kenya	Luhya	[1]
DQ304897	L0a1b2	AM	USA	-	[10]
DQ304898	L0a1b2	AM	USA	-	[10]
EU092963	L0a1b2	AM	USA	-	[5]
KJ185497	L0a1b2	SA	Angola	Kuvale	[4]
KJ185502	L0a1b2	SA	Angola	Kuvale	[4]
KJ185507	L0a1b2	SA	Angola	Kuvale	[4]
KJ185750	L0a1b2	SA	Angola	Nyaneka	[4]
KJ185758	L0a1b2	SA	Angola	Nyaneka	[4]
KJ186008	L0a1b2	SA	Angola	Nyaneka	[4]
EU092746	L0a1b2a	AP / NE	S. Arabia	-	[5]
Howell560	L0a1b2a	AM	USA	-	[11]
KJ185551	L0a1b2a	SA	Zambia	Lozi	[4]
KJ185805	L0a1b2a	SA	Angola	Ovimbundu	[4]
KJ185926	L0a1b2a	SA	Zambia	Makoma	[4]
EU092760	L0a1c1	AS	Iran	-	[5]
EU092945	L0a1c1	EA	Ethiopia	-	[5]
NA19467	L0a1c1	EA	Kenya	Luhya	[1]
KF672811	L0a1c2	CA	Chad	Daza	[3]
NA19440	L0a1c2	EA	Kenya	Luhya	[1]
EU092670	L0a1d	EA	Ethiopia	Jew	[5]
EU092801	L0a1d	AP / NE	Yemen	-	[5]
EU092809	L0a1d	AP / NE	Yemen	-	[5]
EU092810	L0a1d	AP / NE	Yemen	-	[5]
EU092950	L0a1d	EA	Ethiopia	-	[5]
KF672812	L0a1d	EA	Ethiopia	Oromo	[3]
KF672815	L0a1d	EA	Kenya	Turkana	[3]
KF672820	L0a1d	EA	Somalia	-	[3]
EF184602	L0a2	EA	Tanzania	-	[15]
EF184604	L0a2	EA	Tanzania	-	[15]
EF184606	L0a2	EA	Tanzania	-	[15]
EF184607	L0a2	EA	Tanzania	-	[15]
EF184608	L0a2	EA	Tanzania	-	[15]
KF672810	L0a2	EA	Somalia	-	[3]
HM771160	L0a2a1	CA	CAR	Pygmy	[16]
HM771161	L0a2a1	CA	CAR	Pygmy	[16]

Annex 3. *(continued)*

KC622154	L0a2a1a	SA	Namibia	Owambo	[14]
KF672831	L0a2a1a	CA	Niger	Zinder	[3]
KJ185396	L0a2a1a	SA	Zambia	Bemba	[4]
KJ185433	L0a2a1a	SA	Zambia	Luvale	[4]
KJ185722	L0a2a1a	SA	Zambia	Nkoya	[4]
KJ185751	L0a2a1a	SA	Angola	Nyaneka	[4]
KJ185752	L0a2a1a	SA	Angola	Nyaneka	[4]
KJ185881	L0a2a1a	SA	Zambia	Kwangwa	[4]
KJ185882	L0a2a1a	SA	Zambia	Kwangwa	[4]
KJ185883	L0a2a1a	SA	Zambia	Kwangwa	[4]
KJ185424	L0a2a1a1	SA	Zambia	Luchazi	[4]
EU092868	L0a2a1a2	SA	S. Africa	SWB	[5]
EU092911	L0a2a1a2	EA	Kenya	-	[5]
KC622065	L0a2a1a2	SA	Botswana	Tswana	[14]
KF672834	L0a2a1a2	WA	STP	-	[3]
KJ185968	L0a2a1a2	SA	Zambia	Shona	[4]
EU092861	L0a2a1b	SA	S. Africa	SWB	[5]
JX303784	L0a2a1b	SA	Zambia	Tonga	[2]
JX303830	L0a2a1b	SA	Zambia	Kwamashi	[2]
KJ185397	L0a2a1b	SA	Zambia	Bemba	[4]
KJ185461	L0a2a1b	SA	Zambia	Tonga	[4]
KJ185462	L0a2a1b	SA	Zambia	Tonga	[4]
KJ185476	L0a2a1b	SA	Angola	Ganguela	[4]
KJ185480	L0a2a1b	SA	Angola	Ganguela	[4]
KJ185494	L0a2a1b	SA	Angola	Kuvale	[4]
KJ185495	L0a2a1b	SA	Angola	Kuvale	[4]
KJ185496	L0a2a1b	SA	Angola	Kuvale	[4]
KJ185498	L0a2a1b	SA	Angola	Kuvale	[4]
KJ185499	L0a2a1b	SA	Angola	Kuvale	[4]
KJ185500	L0a2a1b	SA	Angola	Kuvale	[4]
KJ185501	L0a2a1b	SA	Angola	Kuvale	[4]
KJ185503	L0a2a1b	SA	Angola	Kuvale	[4]
KJ185505	L0a2a1b	SA	Angola	Kuvale	[4]
KJ185506	L0a2a1b	SA	Angola	Kuvale	[4]
KJ185508	L0a2a1b	SA	Angola	Kuvale	[4]
KJ185543	L0a2a1b	SA	Zambia	Lozi	[4]
KJ185546	L0a2a1b	SA	Zambia	Lozi	[4]
KJ185557	L0a2a1b	SA	Zambia	Tonga	[4]
KJ185653	L0a2a1b	SA	Zambia	Mbunda	[4]
KJ185654	L0a2a1b	SA	Zambia	Mbunda	[4]
KJ185717	L0a2a1b	SA	Zambia	Nkoya	[4]
KJ185720	L0a2a1b	SA	Zambia	Nkoya	[4]
KJ185755	L0a2a1b	SA	Angola	Nyaneka	[4]
KJ185865	L0a2a1b	SA	Zambia	Kwamulonga	[4]
FJ157838	L0a2a2	AS	India	-	[17]
FJ157839	L0a2a2	AS	India	-	[17]
FJ157840	L0a2a2	AS	India	-	[17]
DQ341058	L0a2a2a	AM	Dominica	-	[18]

Annex 3. *(continued)*

EU092701	L0a2a2a	SA	Mozambique	Shangaan	[5]
EU092745	L0a2a2a	AP / NE	S. Arabia	-	[5]
EU092787	L0a2a2a	AP / NE	Oman	-	[5]
EU092925	L0a2a2a	AP / NE	Oman	-	[5]
Howell149	L0a2a2a	AM	USA	-	[11]
JQ705109	L0a2a2a	-	unknown	-	[7]
JX303763	L0a2a2a	SA	Zambia	Totela	[2]
JX303772	L0a2a2a	SA	Zambia	Totela	[2]
JX303778	L0a2a2a	SA	Zambia	Totela	[2]
JX303786	L0a2a2a	SA	Zambia	Tonga	[2]
JX303826	L0a2a2a	SA	Zambia	Kwamashi	[2]
JX303904	L0a2a2a	SA	Zambia	Totela	[2]
KC622056	L0a2a2a	SA	Botswana	Tswana	[14]
KC622064	L0a2a2a	SA	Botswana	Tswana	[14]
KC622187	L0a2a2a	SA	Namibia	Kwanyama	[14]
KF672819	L0a2a2a	SA	Mozambique	-	[3]
KF672824	L0a2a2a	WA	STP	-	[3]
KF672825	L0a2a2a	EA	Somalia	-	[3]
KF672832	L0a2a2a	SA	Mozambique	-	[3]
KF672835	L0a2a2a	EA	Somalia	-	[3]
KJ185394	L0a2a2a	SA	Zambia	Aushi	[4]
KJ185458	L0a2a2a	SA	Zambia	Tokaleya	[4]
KJ185477	L0a2a2a	SA	Angola	Ganguela	[4]
KJ185478	L0a2a2a	SA	Angola	Ganguela	[4]
KJ185479	L0a2a2a	SA	Angola	Ganguela	[4]
KJ185544	L0a2a2a	SA	Zambia	Lozi	[4]
KJ185556	L0a2a2a	SA	Zambia	Lozi	[4]
KJ185657	L0a2a2a	SA	Zambia	Mbunda	[4]
KJ185658	L0a2a2a	SA	Zambia	Mbunda	[4]
KJ185721	L0a2a2a	SA	Zambia	Nkoya	[4]
KJ185749	L0a2a2a	SA	Angola	Nyaneka	[4]
KJ185802	L0a2a2a	SA	Angola	Ovimbundu	[4]
KJ185804	L0a2a2a	SA	Angola	Ovimbundu	[4]
KJ185862	L0a2a2a	SA	Zambia	Tswana	[4]
KJ185970	L0a2a2a	SA	Zambia	Chewa	[4]
NA19312	L0a2a2a	EA	Kenya	Luhya	[1]
NA19328	L0a2a2a	EA	Kenya	Luhya	[1]
NA19402	L0a2a2a	EA	Kenya	Luhya	[1]
NA19904	L0a2a2a	AM	USA	-	[1]
JX303831	L0a2a2a1	SA	Zambia	Kwamashi	[2]
JX303835	L0a2a2a1	SA	Zambia	Kwamashi	[2]
KJ185541	L0a2a2a1	SA	Zambia	Lozi	[4]
KJ185550	L0a2a2a1	SA	Zambia	Lozi	[4]
KJ185553	L0a2a2a1	SA	Zambia	Lozi	[4]
KJ185655	L0a2a2a1	SA	Zambia	Mbunda	[4]
KJ185719	L0a2a2a1	SA	Zambia	Nkoya	[4]
KJ185803	L0a2a2a1	SA	Angola	Ovimbundu	[4]
KJ185946	L0a2a2a1	SA	Zambia	Nyengo	[4]

Annex 3. *(continued)*

KJ185947	L0a2a2a1	SA	Zambia	Nyengo	[4]
NA19713	L0a2a2a1	AM	USA	-	[1]
NA19985	L0a2a2a1	AM	USA	-	[1]
AF346998	L0a2b	CA	Congo	Mbuti Pygmy	[8]
AF346999	L0a2b	CA	Congo	Mbuti Pygmy	[8]
AM711903	L0a2b	CA	Congo	Pygmy	[19]
EU597537	L0a2b	CA	Congo	Pygmy	[20]
HM771189	L0a2b	CA	Congo	Pygmy	[16]
HM771190	L0a2b	CA	Congo	Pygmy	[16]
HM771199	L0a2b	CA	Congo	Pygmy	[16]
HM771202	L0a2b	CA	Congo	Pygmy	[16]
HM771188	L0a2b1	CA	Congo	Pygmy	[16]
HM771200	L0a2b1	CA	Congo	Pygmy	[16]
HM771201	L0a2b1	CA	Congo	Pygmy	[16]
EF556174	L0a2c	EA	Ethiopia	Jew	[21]
KF672813	L0a2c	EA	Somalia	-	[3]
EU092913	L0a2d	EA	Kenya	-	[5]
KJ185399	L0a2d	SA	Zambia	Bemba	[4]
KJ185463	L0a2d	SA	Zambia	Tonga	[4]
EU092900	L0a3	CA	Chad	Sara	[5]
KF672796	L0a3	CA	Cameroon	Bulahay	[3]
EU092906	L0a4	EA	Kenya	-	[5]

References (Annex 3)

1. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA: **An integrated map of genetic variation from 1,092 human genomes.** *Nature* 2012, **491**:56–65.
2. Barbieri C, Butthof A, Bostoen K, Pakendorf B: **Genetic perspectives on the origin of clicks in Bantu languages from southwestern Zambia.** *European Journal of Human Genetics* 2013, **21**:430–6.
3. Rito T, Richards MB, Fernandes V, Alshamali F, Cerny V, Pereira L, Soares P: **The first modern human dispersals across Africa.** *PloS one* 2013, **8**:e80031.
4. Barbieri C, Vicente M, Oliveira S, Bostoen K, Rocha J, Stoneking M, Pakendorf B: **Migration and Interaction in a Contact Zone: mtDNA Variation among Bantu-Speakers in Southern Africa.** *PloS one* 2014, **9**:e99117.
5. Behar DM, Villems R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, Scozzari R, Makkan H, Tzur S, Comas D, Bertranpetit J, Quintana-Murci L, Tyler-Smith C, Wells RS, Rosset S: **The Dawn of Human Matrilineal Diversity**. *The American Journal of Human Genetics* 2008, **82**:1130–1140.
6. Barbieri C, Whitten M, Beyer K, Schreiber H, Li M, Pakendorf B: **Contrasting maternal and paternal histories in the linguistic context of Burkina Faso.** *Molecular Biology and Evolution* 2012, **29**:1213–23.
7. Behar DM, van Oven M, Rosset S, Metspalu M, Loogväli E-L, Silva NM, Kivisild T, Torroni A, Villems R: **A "Copernican" reassessment of the human mitochondrial DNA tree from its root.** *American Journal of Human Genetics* 2012, **90**:675–84.
8. Ingman M, Kaessmann H, Pääbo S, Gyllensten U: **Mitochondrial genome variation and the origin of modern humans.** *Nature* 2000, **408**:708–13.
9. Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC: **Natural selection shaped regional mtDNA variation in humans.** *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**:171–6.

10. Just RS, Diegoli TM, Saunier JL, Irwin JA, Parsons TJ: **Complete mitochondrial genome sequences for 265 African American and U.S. "Hispanic" individuals.** *Forensic Science International Genetics* 2008, **2**:e45–8.
11. Howell N, Elson JL, Turnbull DM, Herrnstadt C: **African Haplogroup L mtDNA sequences show violations of clock-lfike evolution.** *Molecular Biology and Evolution* 2004, **21**:1843–54.
12. Maca-Meyer N, Gonzalez A, Larruga J, Flores C, Cabrera V: **Major genomic mitochondrial lineages delineate early human expansions**. *BMC Genetics* 2001, **2**:13.
13. Kujanová M, Pereira L, Fernandes V, Pereira JB, Cerný V: **Near eastern neolithic genetic input in a small oasis of the Egyptian Western Desert.** *American Journal of Physical Anthropology* 2009, **140**:336–46.
14. Barbieri C, Güldemann T, Naumann C, Gerlach L, Berthold F, Nakagawa H, Mpoloka SW, Stoneking M, Pakendorf B: **Unraveling the complex maternal history of Southern African Khoisan populations.** *American Journal of Physical Anthropology* 2014, **153**:435–48.
15. Gonder MK, Mortensen HM, Reed F a, de Sousa A, Tishkoff S a: **Whole-mtDNA genome sequence analysis of ancient African lineages.** *Molecular Biology and Evolution* 2007, **24**:757–68.
16. Batini C, Lopes J, Behar DM, Calafell F, Jorde LB, van der Veen L, Quintana-Murci L, Spedini G, Destro-Bisol G, Comas D: **Insights into the demographic history of African Pygmies from complete mitochondrial genomes.** *Molecular Biology and Evolution* 2011, **28**:1099–110.
17. Eaaswarkhanth M, Haque I, Ravesh Z, Romero IG, Meganathan PR, Dubey B, Khan FA, Chaubey G, Kivisild T, Tyler-Smith C, Singh L, Thangaraj K: **Traces of sub-Saharan and Middle Eastern lineages in Indian Muslim populations.** *European Journal of Human Genetics : EJHG* 2010, **18**:354–63.
18. Torroni A, Achilli A, Macaulay V, Richards M, Bandelt H-J: **Harvesting the fruit of the human mtDNA tree.** *Trends in Genetics* 2006, **22**:339–45.
19. Arnason U, Gullberg A, Janke A, Kullberg M: **Mitogenomic analyses of caniform relationships.** *Molecular Phylogenetics and Evolution* 2007, **45**:863–74.
20. Hartmann A, Thieme M, Nanduri LK, Stempfl T, Moehle C, Kivisild T, Oefner PJ: **Validation of microarray-based resequencing of 93 worldwide mitochondrial genomes.** *Human Mutation* 2009, **30**:115–22.
21. Behar DM, Metspalu E, Kivisild T, Rosset S, Tzur S, Hadid Y, Yudkovsky G, Rosengarten D, Pereira L, Amorim A, Kutuev I, Gurwitz D, Bonne-Tamir B, Villems R, Skorecki K: **Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora.** *PloS one* 2008, **3**:e2062.

Annex 4. GPS coordinates in WGS-84 (World Geodetic System, 1984) of capital cities used as reference points to construct L2 frequency distribution and references for the 12880 HVRI sequences used to assess L2 frequency.

Codes for regions as in Table 1. Abbreviations: BF – Burkina Faso, CV – Cape Verde, Eq. Guinea – Equatorial Guinea, STP – São Tomé and Príncipe, S. Leone – Sierra Leone, W. Sahara – Western Sahara.

Region	Country	Capital	GPS Coordinates (WGS 84)	References
WA	BF	Ouagadougou	12°21′26″ N 1°32′07″ W	[1–3]
	CV	Praia	14°55’15" N 23° 30’30" W	[4]
	Gambia	Banjul	13°27′11″ N 16°34′39″ W	[5]
	Ghana	Accra	5°33′00″ N 0°12′00″ W	[6]
	Mali	Bamako	12°39′00″ N 8°00′00″ W	[1, 3, 7, 8]
	STP	São Tomé	0°20′10″ N 6°40′53″ E	[9, 10]
	Senegal	Dakar	14°41′34″ N 17°26′48″ W	[2, 11, 12]
	S.Leone	Freetown	8°29′4″ N 13°14′04″ W	[5, 13]
CA	Cameroon	Yaoundé	3°52′00‘‘ N 11°31′00‘‘ E	[3, 6, 14]
	Chad	N’Djamena	12°06′47″ N 15°02′57″ E	[3, 11]
	Congo	Brazzaville	4°16′04″ S 15°17′31″ E	[11]
	Eq. Guinea	Malabo	3°45′00‘‘ N 8°46′59″ E	[9]
	Niger	Niamey	13°31′17″ N 02°06′19″ E	[1, 3, 11]
	Nigeria	Abuja	9°04′00″ N 7°29′00″ E	[3, 5, 6, 11]
	W. Sahara	El Aaiún	27°09′13″ N 13°12′12″ W	[12, 15]
EA	Ethiopia	Addis Ababa	9°01′48″ N 38°44′24″ E	[16–19]
	Kenya	Nairobi	1°17′31‘‘ S 36°49′19’’ E	[5, 11, 20, 21]
	Rwanda	Kigali	1°56′38″ S 30°03′34″ E	[22]
	Somalia	Mogadishu	02°02′ N 45°21′ E	[11, 19]
	Sudan	Khartoum	15°32′47‘‘ N 32°32′00’’ E	[19, 23]
	Tanzania	Dodoma	06° 10’23" S 35° 44’31" E	[24]
SA	Angola	Luanda	8°50′18″ S 13°14′04″ E	[25–27]
	Botswana	Gaborone	24°39′29″ S 25°54′44″ E	[28]
	Cabinda	Cabinda	5°33′36″ S 12°11′24″ E	[29]
	Madagascar	Antananarivo	18°54′51‘‘ S 47°31′51‘‘ E	[30]
	Mozambique	Maputo	25°57′55″ S 32°35′21″ E	[31, 32]
	Namibia	Windhoek	22°34′12″ S 17°05′01″ E	[28]
	South Africa	Pretoria	25°44′46″ S 28°11′17″ E	[33, 34]
	Zambia	Lusaka	15°24′29″ S 28°17′10″ E	[35, 36]
	Zimbabwe	Harare	17°51′50″ S 31°01′47″ E	[22]
NA	Algeria	Algiers	36º45′08″ N 3°02′31″ E	[15]
	Egypt	Cairo	30°03′29″ N 31°13′44″ E	[37, 38]
	Libya	Tripoli	32°54′08″ N 13°11′09″ E	[39]
	Mauritania	Nouakchott	18°06′01″ N 15°56′59″ W	[8, 12]
	Morocco	Rabat	34°00′47″ N 6°49′57″ W	[12, 37, 40]
	Tunisia	Tunis	36°49′08″ N 10°09′56″ E	[15, 41–44]

References (Annex 4)

1. Pereira L, Cerny V, Cerezo M, Silva NM, Hajek M, Vasikova A, Kujanova M, Brdicka R, Salas A: **Linking the sub-Saharan and West Eurasian gene pools: maternal and paternal heritage of the Tuareg nomads from the African Sahel**. *European Journal of Human Genetics* 2010, **18**:915–923.
2. Barbieri C, Whitten M, Beyer K, Schreiber H, Li M, Pakendorf B: **Contrasting maternal and paternal histories in the linguistic context of Burkina Faso.** *Molecular Biology and Evolution* 2012, **29**:1213–23.
3. Černý V, Pereira L, Musilová E, Kujanová M, Vašíková A, Blasi P, Garofalo L, Soares P, Diallo I, Brdička R, Novelletto A: **Genetic Structure of Pastoral and Farmer Populations in the African Sahel**. *Molecular Biology and Evolution* 2011, **28**:2491–2500.
4. Brehm A, Pereira L, Bandelt HJ, Prata MJ, Amorim A: **Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade**. *Annals of Human Genetics* 2002, **66**:49–60.
5. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA: **An integrated map of genetic variation from 1,092 human genomes.** *Nature* 2012, **491**:56–65.
6. Veeramah KR, Connell BA, Ansari Pour N, Powell A, Plaster CA, Zeitlyn D, Mendell NR, Weale ME, Bradman N, Thomas MG: **Little genetic differentiation as assessed by uniparental markers in the presence of substantial language variation in peoples of the Cross River region of Nigeria.** *BMC Evolutionary Biology* 2010, **10**:92.
7. Ely B, Wilson JL, Jackson F, Jackson BA: **African-American mitochondrial DNAs often match mtDNAs found in multiple African ethnic groups.** *BMC Biology* 2006, **4**:34.
8. González AM, Cabrera VM, Larruga JM, Tounkara A, Noumsi G, Thomas BN, Moulds JM: **Mitochondrial DNA variation in Mauritania and Mali and their genetic relationship to other Western Africa populations.** *Annals of Human Genetics* 2006, **70**(5):631–57.
9. Mateu E, Comas D, Calafell F, Pérez-Lezaun A, Abade A, Bertranpetit J: **A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and São Tomé, Gulf of Guinea.** *Annals of Human Genetics* 1997, **61**(6):507–18.
10. Trovoada MJ, Pereira L, Gusmao L, Abade A, Amorim A, Prata MJ: **Pattern of mtDNA Variation in Three Populations from Sao Tome e Principe**. *Annals of Human Genetics* 2004, **68**:40–54.
11. Watson E, Forster P, Richards M, Bandelt H: **Mitochondrial footprints of human expansions in Africa**. *American Journal of Human Genetics* 1997, **61**(3):691–704.
12. Rando JC, Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM, Bandelt H-J: **Mitochondrial DNA analysis of Northwest African populations reveals genetic exchanges with European, Near-Eastern, and sub-Saharan populations**. *Annals of Human Genetics* 1998, **62**:531–550.
13. Jackson BA, Wilson JL, Kirbah S, Sidney SS, Rosenberger J, Bassie L, Alie JAD, McLean DC, Garvey WT, Ely B: **Mitochondrial DNA genetic diversity among four ethnic groups in Sierra Leone.** *American Journal of Physical Anthropology* 2005, **128**:156–63.
14. Quintana-Murci L, Quach H, Harmant C, Luca F, Massonnet B, Patin E, Sica L, Mouguiama-Daouda P, Comas D, Tzur S, Balanovsky O, Kidd KK, Kidd JR, van der Veen L, Hombert J-M, Gessain A, Verdu P, Froment A, Bahuchet S, Heyer E, Dausset J, Salas A, Behar DM: **Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers.** *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**:1596–601.
15. Plaza S, Calafell F, Helal A, Bouzerna N, Lefranc G, Bertranpetit J, Comas D: **Joining the Pillars of Hercules: mtDNA Sequences Show Multidirectional Gene Flow in the Western Mediterranean**. *Annals of Human Genetics* 2003, **67**:312–328.

16. Poloni ES, Naciri Y, Bucho R, Niba R, Kervaire B, Excoffier L, Langaney A, Sanchez-Mazas A: **Genetic evidence for complexity in ethnic differentiation and history in East Africa.** *Annals of Human Genetics* 2009, **73**(6):582–600.
17. Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, Villems R: **Ethiopian Mitochondrial DNA Heritage: Tracking Gene Flow Across and Around the Gate of Tears**. *The American Journal of Human Genetics* 2004, **75**:752–770.
18. Non AL, Al-Meeri A, Raaum RL, Sanchez LF, Mulligan CJ: **Mitochondrial DNA reveals distinct evolutionary histories for Jewish populations in Yemen and Ethiopia.** *American Journal of Physical Anthropology* 2011, **144**:1–10.
19. Soares P, Alshamali F, Pereira JB, Fernandes V, Silva NM, Afonso C, Costa MD, Musilová E, Macaulay V, Richards MB, Cerny V, Pereira L: **The Expansion of mtDNA Haplogroup L3 within and out of Africa.** *Molecular Biology and Evolution* 2012, **29**:915–27.
20. Brandstätter A, Peterson CT, Irwin JA, Mpoke S, Koech DK, Parson W, Parsons TJ: **Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database.** *International Journal of Legal Medicine* 2004, **118**:294–306.
21. Boattini A, Castrì L, Sarno S, Useli A, Cioffi M, Sazzini M, Garagnani P, De Fanti S, Pettener D, Luiselli D: **mtDNA variation in East Africa unravels the history of Afro-Asiatic groups.** *American Journal of Physical Anthropology* 2013, **150**:375–85.
22. Castrì L, Tofanelli S, Garagnani P, Bini C, Fosella X, Pelotti S, Paoli G, Pettener D, Luiselli D: **mtDNA variability in two Bantu-speaking populations (Shona and Hutu) from Eastern Africa: implications for peopling and migration patterns in sub-Saharan Africa.** *American Journal of Physical Anthropology* 2009, **140**:302–11.
23. Krings M, Salem AH, Bauer K, Geisert H, Malek AK, Chaix L, Simon C, Welsby D, Di Rienzo A, Utermann G, Sajantila A, Pääbo S, Stoneking M: **mtDNA Analysis of Nile River Valley Populations: A Genetic Corridor or a Barrier to Migration?** *The American Journal of Human Genetics* 1999, **64**:1166–1176.
24. Knight A, Underhill P, Mortensen H: **African Y chromosome and mtDNA divergence provides insight into the history of click languages**. *Current Biology* 2003, **13**:464–473.
25. Coelho M, Sequeira F, Luiselli D, Beleza S, Rocha J: **On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola**. *BMC Evolutionary Biology* 2009, **9**:80.
26. Plaza S, Salas A, Calafell F, Corte-Real F, Bertranpetit J, Carracedo A, Comas D: **Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola.** *Human Genetics* 2004, **115**:439–47.
27. Barbieri C, Vicente M, Oliveira S, Bostoen K, Rocha J, Stoneking M, Pakendorf B: **Migration and Interaction in a Contact Zone: mtDNA Variation among Bantu-Speakers in Southern Africa.** *PloS one* 2014, **9**:e99117.
28. Barbieri C, Güldemann T, Naumann C, Gerlach L, Berthold F, Nakagawa H, Mpoloka SW, Stoneking M, Pakendorf B: **Unraveling the complex maternal history of Southern African Khoisan populations.** *The American Journal of Physical Anthropology* 2014, **153**:435–48.
29. Beleza S, Gusmão L, Amorim A, Carracedo A, Salas A: **The genetic legacy of western Bantu migrations.** *Human Genetics* 2005, **117**:366–75.
30. Tofanelli S, Bertoncini S, Castrì L, Luiselli D, Calafell F, Donati G, Paoli G: **On the origins and admixture of Malagasy: new evidence from high-resolution analyses of paternal and maternal lineages.** *Molecular Biology and Evolution* 2009, **26**:2109–24.
31. Salas A, Richards M, De la Fe T, Lareu M-V, Sobrino B, Sánchez-Diz P, Macaulay V, Carracedo A: **The making of the African mtDNA landscape.** *The American Journal of Human Genetics* 2002, **71**:1082–111.

32. Pereira L, Macaulay V, Torroni A, Scozzari R, Pavia Z, Ferrata V: **Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade**. *Annals of Human Genetics* 2001, 65(5):439–458.
33. Quintana-Murci L, Harmant C, Quach H, Balanovsky O, Zaporozhchenko V, Bormans C, van Helden PD, Hoal EG, Behar DM: **Strong maternal Khoisan contribution to the South African coloured population: a case of gender-biased admixture.** *The American Journal of Human Genetics* 2010, **86**:611–20.
34. Chen YS, Olckers A, Schurr TG, Kogelnik AM, Huoponen K, Wallace DC: **mtDNA variation in the South African Kung and Khwe-and their genetic relationships to other African populations.** *The American Journal of Human Genetics* 2000, **66**:1362–83.
35. De Filippo C, Heyn P, Barham L, Stoneking M, Pakendorf B: **Genetic perspectives on forager-farmer interaction in the Luangwa valley of Zambia.** *American Journal of Physical Anthropology* 2010, **141**:382–94.
36. Barbieri C, Butthof A, Bostoen K, Pakendorf B: **Genetic perspectives on the origin of clicks in Bantu languages from southwestern Zambia.** *European Journal of Human Genetics* 2013, **21**:430–6.
37. Coudray C, Olivieri A, Achilli A, Pala M, Melhaoui M, Cherkaoui M, El-Chennawi F, Kossmann M, Torroni A, Dugoujon JM: **The complex and diversified mitochondrial gene pool of Berber populations.** *Annals of Human Genetics* 2009, **73**:196–214.
38. Saunier JL, Irwin JA, Strouss KM, Ragab H, Sturk KA, Parsons TJ: **Mitochondrial control region sequences from an Egyptian population sample.** *Forensic Science International Genetics* 2009, **3**:e97–103.
39. Ottoni C, Martínez-Labarga C, Loogväli E-L, Pennarun E, Achilli A, De Angelis F, Trucchi E, Contini I, Biondi G, Rickards O: **First genetic insight into Libyan Tuaregs: a maternal perspective.** *Annals of Human Genetics* 2009, **73**(Pt 4):438–48.
40. Rhouda T, Martínez-Redondo D, Gómez-Durán A, Elmtili N, Idaomar M, Díez-Sánchez C, Montoya J, López-Pérez MJ, Ruiz-Pesini E: **Moroccan mitochondrial genetic background suggests prehistoric human migrations across the Gibraltar Strait.** *Mitochondrion* 2009, **9**:402–7.
41. Fadhlaoui-Zid K, Plaza S, Calafell F, Ben Amor M, Comas D, Bennamar El gaaied A: **Mitochondrial DNA heterogeneity in Tunisian Berbers.** *Annals of Human Genetics* 2004, **68**(3):222–33.
42. Loueslati BY, Cherni L, Khodjet-Elkhil H, Ennafaa H, Pereira L, Amorim A, Ben Ayed F, Ben Ammar Elgaaied A: **Islands inside an island: reproductive isolates on Jerba island.** *American Journal of Human Biology* 2006, **18**:149–53.
43. Turchi C, Buscemi L, Giacchino E, Onofri V, Fendt L, Parson W, Tagliabracci A: **Polymorphisms of mtDNA control region in Tunisian and Moroccan populations: an enrichment of forensic mtDNA databases with Northern Africa data.** *Forensic Science International Genetics* 2009, **3**:166–72.
44. Cherni L, Fernandes V, Pereira JB, Costa MD, Goios A, FRIGI S, Yacoubi-Loueslati B, Amor M Ben, Slam A, Amorim A, Ben Ammar el Gaaied A, Pereira L: **Post-Last Glacial Maximum Expansion From Iberia to North Africa Revealed by Fine Characterization of mtDNA H Haplogroup in Tunisia**. *American Journal of Physical Anthropology* , **139**:253–260.

Annex 5. Information on the populations used to assess pairwise genetic distances and to compute the MDS plot. References for the 4880 HVRI sequences used.

Codes for regions as in Table 1. Abbreviations: Eq. Guinea – Equatorial Guinea, STP – São Tomé and Príncipe.

Region	Country	Population	Code	N	Language group	Additional information	Reference
EA	Ethiopia	Daasanach	Daa	49	Cushitic	Agropastoralist	[1]
		Dawro-Konta	Daw	137	Omotic	Agropastoralist	[2]
		Nyangatom	Nya	112	Nilotic	Agropastoralist	[1]
		General population	Eth	77	-	-	[3]
	Kenya	El Molo	Elm	52	Cushitic	Fishing	[2]
		Luo	Luo	49	Nilotic	Agropastoralist, fishing	[2]
		Luhya	LWK	120	Bantu	Agriculturalist	[4]
		Maasai	Maa	81	Nilotic	Seminomadic pastoralist	[2]
		Turkana	Turk	37	Nilotic	Seminomadic pastoralist	[5]
	Rwanda	Hutu	Hutu	42	Bantu	-	[6]
	Somalia	General population	Som	177	-	-	[3, 5]
	Sudan	General population	Sud	178	-	-	[3, 7]
	Tanzania	Burunge	Buru	38	Cushitic	Agriculturalist	[8]
		Datog	Dat	31	Nilotic	Pastoralist, Agriculturalist	[8]
WA	Burkina Faso	Various groups	BF	291	-	-	[9]
	Cape Verde	General population	CV	295	-	-	[10]
	Gambia	Western Division	GWD	113			[4]
	Mali	Bambara	Mali	158	Mande	-	[11, 12]
		Malinke			Mande	-	
	Senegal	Mandenka	Man	131	Mande	-	[5, 9]
		Wolof	Wol	61	Wolof	-	[13, 14]
	Sierra Leone	Mende	MSL	85	Mel	-	[4]
CA	Eq. Guinea	Bioko	Bio	45	Bantu	-	[15]
	Cameroon	Ngumba	Ngu	88	Bantu	-	[16]
	Gabon	Akele	Ake	48	Bantu	-	[16]
		Ateke	Ate	54	Bantu	-	
		Benga	Beng	50	Bantu	-	
		Duma	Duma	47	Bantu	-	
		Eshira	Eshi	40	Bantu	-	
		Eviya	Evi	38	Bantu	-	
		Fang	Fang	66	Bantu	-	
		Galoa	Gal	51	Bantu	-	
		Kota	Kota	56	Bantu	-	
		Makina	Mak	45	Bantu	-	

Annex 5. *(continued)*

		Mitsogo	Mits	64	Bantu	-	
		Nzebi	Nze	63	Bantu	-	
		Punu	Punu	52	Bantu	-	
		Shake	Sha	51	Bantu	-	
	Nigeria	Esan	ESN	99	Edoid	-	[4]
		Yoruba	Yor	193	Yoruboid	-	[4, 5, 9]
	STP	General population	STP	54	-	-	[15]
SA	Angola	Mbundu	Ang	466	Bantu	-	[17]
		West-Savanna			Bantu	-	[18]
		Cabinda (Fiote)	Cab	109	Bantu	-	[19]
		Kuvale	Kuv	55	Bantu	Seminomadic pastoralists	[20]
		Nyaneka	Nyane	59	Bantu	Agriculturalist	[20]
	Mozambique	General population	Moz	187	-	-	[21, 22]
	Zambia	Bisa	Bisa	42	Bantu	-	[23]
		Fwe	Fwe	33	Bantu	Agropastoralist	[24]
		Kunda	Kunda	36	Bantu	-	[23]
		Kwamashi	Kwam	35	Bantu	Agriculturalist	[20]
		Kwangwa	Kwang	35	Bantu	-	[20]
		Lozi	Lozi	110	Bantu	Agriculturalist	[20]
		Mbunda	Mbun	64	Bantu	Agriculturalist	[20]
		Nkoya	Nkoy	32	Bantu	Agriculturalist	[20]
		Tonga	Ton	37	Bantu	Agriculturalist	[24]
	Zimbabwe	Shona	Sho	62	Bantu	-	[6]

References (Annex 5)

1. Poloni ES, Naciri Y, Bucho R, Niba R, Kervaire B, Excoffier L, Langaney A, Sanchez-Mazas A: **Genetic evidence for complexity in ethnic differentiation and history in East Africa.** *Annals of Human Genetics* 2009, **73**(6):582–600.
2. Boattini A, Castrì L, Sarno S, Useli A, Cioffi M, Sazzini M, Garagnani P, De Fanti S, Pettener D, Luiselli D: **mtDNA variation in East Africa unravels the history of Afro-Asiatic groups.** *American Journal of Physical Anthropology* 2013, **150**:375–85.
3. Soares P, Alshamali F, Pereira JB, Fernandes V, Silva NM, Afonso C, Costa MD, Musilová E, Macaulay V, Richards MB, Cerny V, Pereira L: **The Expansion of mtDNA Haplogroup L3 within and out of Africa.** *Molecular Biology and Evolution* 2012, **29**:915–27.
4. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA: **An integrated map of genetic variation from 1,092 human genomes.** *Nature* 2012, **491**:56–65.
5. Watson E, Forster P, Richards M, Bandelt H: **Mitochondrial footprints of human expansions in Africa**. *The American Journal of Human Genetics* 1997, **61**(3):691–704.
6. Castrì L, Tofanelli S, Garagnani P, Bini C, Fosella X, Pelotti S, Paoli G, Pettener D, Luiselli D: **mtDNA variability in two Bantu-speaking populations (Shona and Hutu) from Eastern Africa: implications for peopling**

and migration patterns in sub-Saharan Africa. *American Journal of Physical Anthropology* 2009, **140**:302–11.

7. Krings M, Salem AH, Bauer K, Geisert H, Malek AK, Chaix L, Simon C, Welsby D, Di Rienzo A, Utermann G, Sajantila A, Pääbo S, Stoneking M: **mtDNA Analysis of Nile River Valley Populations: A Genetic Corridor or a Barrier to Migration?** *The American Journal of Human Genetics* 1999, **64**:1166–1176.
8. Tishkoff S a, Gonder MK, Henn BM, Mortensen H, Knight A, Gignoux C, Fernandopulle N, Lema G, Nyambo TB, Ramakrishnan U, Reed F a, Mountain JL: **History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation.** *Molecular Biology and Evolution* 2007, **24**:2180–95.
9. Barbieri C, Whitten M, Beyer K, Schreiber H, Li M, Pakendorf B: **Contrasting maternal and paternal histories in the linguistic context of Burkina Faso.** *Molecular Biology and Evolution* 2012, **29**:1213–23.
10. Brehm A, Pereira L, Bandelt HJ, Prata MJ, Amorim A: **Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade**. *Annals of Human Genetics* 2002, **66**:49–60.
11. Ely B, Wilson JL, Jackson F, Jackson BA: **African-American mitochondrial DNAs often match mtDNAs found in multiple African ethnic groups.** *BMC Biology* 2006, **4**:34.
12. González AM, Cabrera VM, Larruga JM, Tounkara A, Noumsi G, Thomas BN, Moulds JM: **Mitochondrial DNA variation in Mauritania and Mali and their genetic relationship to other Western Africa populations.** *Annals of Human Genetics* 2006, **70**(5):631–57.
13. Rando JC, Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM, Bandelt H-J: **Mitochondrial DNA analysis of Northwest African populations reveals genetic exchanges with European, Near-Eastern, and sub-Saharan populations**. *Annals of Human Genetics* 1998, **62**:531–550.
14. Stefflova K, Dulik MC, Pai AA, Walker AH, Zeigler-Johnson CM, Gueye SM, Schurr TG, Rebbeck TR: **Evaluation of group genetic ancestry of populations from Philadelphia and Dakar in the context of sex-biased admixture in the Americas.** *PloS one* 2009, **4**:e7842.
15. Mateu E, Comas D, Calafell F, Pérez-Lezaun A, Abade A, Bertranpetit J: **A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and São Tomé, Gulf of Guinea.** *Annals of Human Genetics* 1997, **61**(6):507–18.
16. Quintana-Murci L, Quach H, Harmant C, Luca F, Massonnet B, Patin E, Sica L, Mouguiama-Daouda P, Comas D, Tzur S, Balanovsky O, Kidd KK, Kidd JR, van der Veen L, Hombert J-M, Gessain A, Verdu P, Froment A, Bahuchet S, Heyer E, Dausset J, Salas A, Behar DM: **Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers.** *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**:1596–601.
17. Plaza S, Salas A, Calafell F, Corte-Real F, Bertranpetit J, Carracedo A, Comas D: **Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola.** *Human Genetics* 2004, **115**:439–47.
18. Coelho M, Sequeira F, Luiselli D, Beleza S, Rocha J: **On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola**. *BMC Evolutionary Biology* 2009, **9**:80.
19. Beleza S, Gusmão L, Amorim A, Carracedo A, Salas A: **The genetic legacy of western Bantu migrations.** *Human Genetics* 2005, **117**:366–75.
20. Barbieri C, Güldemann T, Naumann C, Gerlach L, Berthold F, Nakagawa H, Mpoloka SW, Stoneking M, Pakendorf B: **Unraveling the complex maternal history of Southern African Khoisan populations.** *American Journal of Physical Anthropology* 2014, **153**:435–48.
21. Salas A, Richards M, De la Fe T, Lareu M-V, Sobrino B, Sánchez-Diz P, Macaulay V, Carracedo A: **The making of the African mtDNA landscape.** *American Journal of Human Genetics* 2002, **71**:1082–111.

22. Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ, Amorim A: **Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade**. *Annals of Human Genetics* 2001, **65**:439–458.

23. De Filippo C, Heyn P, Barham L, Stoneking M, Pakendorf B: **Genetic perspectives on forager-farmer interaction in the Luangwa valley of Zambia.** *American Journal of Physical Anthropology* 2010, **141**:382–94.

24. Barbieri C, Butthof A, Bostoen K, Pakendorf B: **Genetic perspectives on the origin of clicks in Bantu languages from southwestern Zambia.** *European Journal of Human Genetics* 2013, **21**:430–6.

Annex 6. Haplogroup L2 frequency in 39 African populations.

Geographic points considered to construct frequency maps and references for population analysis in Annex 4.

Country/Group	L2 (total)		L2a		L2b		L2d		L2e		L2*		N (total)
	N	F	N	F	N	F	N	F	N	F	N	F	
WA	**877**	**0.371**	**521**	**0.221**	**95**	**0.040**	**29**	**0.012**	**12**	**0.005**	**220**	**0.093**	**2362**
Burkina Faso	182	0.425	116	0.271	24	0.056	6	0.014	5	0.012	31	0.072	428
Cape Verde	121	0.414	60	0.205	12	0.041	0	0.000	0	0.000	49	0.168	292
Gambia	48	0.425	18	0.159	8	0.071	2	0.018	1	0.009	19	0.168	113
Ghana	95	0.399	74	0.311	9	0.038	2	0.008	0	0.000	10	0.042	238
Mali	142	0.300	92	0.194	15	0.032	4	0.008	0	0.000	31	0.065	474
STP	30	0.196	19	0.124	3	0.020	2	0.013	1	0.007	5	0.033	153
Senegal	1	0.374	64	0.212	17	0.056	0	0.000	4	0.013	28	0.093	302
Sierra Leone	146	0.403	78	0.215	7	0.019	13	0.036	1	0.003	47	0.130	362
CA	**907**	**0.218**	**638**	**0.154**	**119**	**0.029**	**32**	**0.008**	**19**	**0.005**	**99**	**0.024**	**4154**
Cameroon	122	0.143	**93**	0.109	15	0.018	7	0.008	3	0.004	4	0.005	851
Chad	35	0.248	25	0.177	7	0.050	0	0.000	3	0.021	0	0.000	141
Eq. Guinea	11	0.244	9	0.200	0	0.000	2	0.044	0	0.000	0	0.000	45
Gabon	165	0.174	118	0.125	15	0.016	13	0.014	0	0.000	19	0.020	946
Niger	57	0.291	22	0.112	10	0.051	0	0.000	7	0.036	18	0.092	196
Nigeria	450	0.263	315	0.184	70	0.041	10	0.006	6	0.004	49	0.029	1714
Western Sahara	13	0.155	3	0.036	1	0.012	0	0.000	0	0.000	9	0.107	84
EA	**297**	**0.151**	**252**	**0.128**	**21**	**0.011**	**4**	**0.002**	**8**	**0.004**	**12**	**0.006**	**1966**
Ethiopia	92	0.127	76	0.105	11	0.015	0	0.000	3	0.004	2	0.003	726
Kenya	83	0.113	72	0.098	5	0.007	1	0.001	4	0.005	1	0.001	734
Rwanda	5	0.119	3	0.071	1	0.024	1	0.024	0	0.000	0	0.000	42
Somalia	34	0.183	27	0.145	0	0.000	0	0.000	0	0.000	7	0.038	186
Sudan	63	0.354	54	0.303	4	0.022	2	0.011	1	0.006	2	0.011	178
Tanzania	20	0.200	20	0.200	0	0.000	0	0.000	0	0.000	0	0.000	100
SA	**498**	**0.158**	**395**	**0.125**	**48**	**0.015**	**6**	**0.002**	**5**	**0.002**	**44**	**0.014**	**3160**
Angola	109	0.181	68	0.113	16	0.027	1	0.002	0	0.000	24	0.040	601
Botswana	27	0.067	25	0.062	1	0.002	0	0.000	0	0.000	1	0.002	402
Cabinda	14	0.128	5	0.046	7	0.064	0	0.000	0	0.000	2	0.018	109
Madagascar	19	0.143	15	0.113	1	0.008	0	0.000	0	0.000	3	0.023	133
Mozambique	149	0.358	137	0.329	6	0.014	3	0.007	0	0.000	3	0.007	416
Namibia	6	0.023	4	0.015	2	0.008	0	0.000	0	0.000	0	0.000	262
South Africa	34	0.053	29	0.046	4	0.006	0	0.000	0	0.000	1	0.002	637
Zambia	126	0.233	100	0.185	10	0.018	2	0.004	5	0.009	9	0.017	541
Zimbabwe	14	0.237	12	0.203	1	0.017	0	0.000	0	0.000	1	0.017	59
NA	**123**	**0.054**	**93**	**0,041**	**14**	**0.006**	**10**	**0.004**	**6**	**0.003**	**0**	**0.000**	**2268**
Algeria	3	0.064	3	0.064	0	0.000	0	0.000	0	0.000	0	0.000	47
Egypt	14	0.039	13	0.037	1	0.003	0	0.000	0	0.000	0	0.000	355
Libya	37	0.103	28	0.078	7	0.020	0	0.000	0	0.000	2	0.006	358
Mauritania	14	0.149	8	0.085	1	0.011	0	0.000	0	0.000	5	0.053	94
Morocco	9	0.010	6	0.007	2	0.002	1	0.001	0	0.000	0	0.000	863
Tunisia	46	0.083	35	0.064	3	0.005	5	0.009	0	0.000	3	0.005	551